

Faktorenanalyse und Clusteranalyse: Möglichkeiten des Einsatzes multivariater Verfahren in der Analyse des Verhältnisses von Stadt und Land im östlichen Münsterland im frühen 19. Jahrhundert

Schwippe, Heinrich Johannes

Veröffentlichungsversion / Published Version

Sammelwerksbeitrag / collection article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Schwippe, H. J. (1979). Faktorenanalyse und Clusteranalyse: Möglichkeiten des Einsatzes multivariater Verfahren in der Analyse des Verhältnisses von Stadt und Land im östlichen Münsterland im frühen 19. Jahrhundert. In W. H. Schröder (Hrsg.), *Moderne Stadtgeschichte* (S. 112-144). Stuttgart: Klett-Cotta. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-327844>

Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Faktorenanalyse und Clusteranalyse

Möglichkeiten des Einsatzes multivariater Verfahren in der Analyse des Verhältnisses von Stadt und Land im östlichen Münsterland im frühen 19. Jahrhundert

1.1 Problemstellung

Seit dem 18. Jahrhundert ist die räumliche Ordnung in vielen deutschen Territorialstaaten durch eine zunehmende Ausweitung der gewerblichen Produktion auch auf dem Lande gekennzeichnet. In einem fortschreitenden Dekonzentrationsprozeß bilden sich außerhalb der Städte auf dem flachen Lande Gebiete mit erheblichen Gewinnen im Gewerbesektor¹. Das seit dem hohen Mittelalter bestehende Prinzip der Arbeitsteilung zwischen Stadt und Land, wobei das Land vor allem der Erzeugung von agrarischen Produkten gedient hat und die Stadt der bevorzugte Standort gewerblicher Produktionsstätten und Dienstleistungen gewesen ist, ist bis zum Ende des 18. Jahrhunderts, allerdings regional unterschiedlich, weitgehend aufgehoben². Dieser Prozeß der Durchdringung des flachen Landes mit gewerblichen Produktionsstätten beinhaltet eine Ausweitung und Differenzierung des Angebotes gewerblicher Güter auf dem Lande und ist damit deutlich von dem, auch zeitlich früher liegenden Prozeß der Ausbildung ländlicher exportorientierter, im allgemeinen monostrukturierter Produktionsgebiete³ zu unterscheiden.

Aufgrund der Entwicklung von Stadt- und Landgewerbe seit dem 18. Jahrhundert ist die in den meisten deutschen Staaten noch fortbestehende rechtliche Sonderstellung der Stadt als Standort von Gewerbe und Dienstleistung also weitgehend

¹ Zum Verhältnis von Stadt- und Landgewerbe vgl. allgemein: Schremmer, E., *Standortausweitung der Warenproduktion im langfristigen Wirtschaftswachstum, Zur Stadt-Land-Arbeitsteilung im Gewerbe des 18. Jahrhunderts*, in: VSWG, 59 (1972), S. 1–40; Kaufhold, K.H., *Umfang und Gliederung des deutschen Handwerks um 1800*, in: Abel, W. u. Mitarb. (Hrsg.), *Handwerksgeschichte in neuerer Sicht*, Göttingen 1970, S. 26–64. Zum Umfang des Landhandwerks im Münsterland vgl.: Nübel, O., *Das Landhandwerk des Münsterlandes um die Wende des 19. Jahrhunderts*, Münster 1913.

² Bezüglich des Ausmaßes der Ausweitung gewerblicher Tätigkeit auf dem Lande ist davon auszugehen, daß ein Ost-West-Gefälle bestanden hat; nach Kaufhold, *Umfang und Gliederung*, dürfte eine wesentliche Ursache für diese Erscheinung die unterschiedliche Agrarstruktur westlich und östlich der Elbe sein.

³ Beispiele für diese exportorientierten *Gewerbelandschaften* mit überwiegend ländlich dispersen Standorten sind für das 18. Jahrhundert die Leinwandgebiete Ostmitteldeutschlands und Minden-Ravensberg.

aufgehoben⁴. Zur Erklärung der Territorialisierung des Gewerbes können eine Vielzahl von Aspekten herangezogen werden; sie lassen sich nach Schremmer zu drei Komplexen zusammenfassen:

„1. das Wachstum der Bevölkerung, 2. sozialökonomische Bedingungen, dazu gehören a) die grund-, gerichts- und landesherrliche Wirtschafts- und Gesellschaftsordnung, b) das Kosten-, Preis- und Einkommensgefüge, c) die Verhaltensweisen der Personen und Personengruppen innerhalb der gegebenen Ordnung und das Ziel-Mittel-Verhältnis ihrer Politik, . . . 3. das angewandte technische und agrarische Wissen . . . , vor allem die Verbreitung des Verlags und der Manufaktur sowie die Veränderungen im Bereich Transport, Verkehr und Kommunikation“⁵.

Gerade weil aufgrund des Bedeutungsverlustes der klassischen Standortfaktoren, u. a. durch Veränderung der territorialen Wirtschaftsordnungen seit dem 18. Jahrhundert eine Ausweitung des Gewerbes auf dem Lande stattfindet, liegt die Vermutung nahe, daß die klassische Gliederung in Stadt und Land durch die stärker mit Gewerbe besetzten Bereiche auf dem Lande modifiziert wird, denn erfahrungsgemäß zieht die Lokalisierung von Gewerbe weitere Strukturveränderungen nach sich: Bevölkerungsverdichtung, höherer Lebensstandard, Ausweitung des tertiären Sektors usw.

⁴ Die Aufhebung des Gegensatzes zwischen Stadt und Land und die Einführung der Gewerbefreiheit, die Abschaffung der traditionellen ständischen Herrschafts- und Wirtschaftsverfassung erfolgt in Preußen im Rahmen der sog. Reformgesetzgebung (Bauernbefreiung, Städteordnung, Gewerbeordnung, . . .) 1807 ff. Für den westfälischen Raum bringt die französische Besetzung bzw. die Eingliederung in die von Frankreich abhängigen Staaten Großherzogtum Berg und Königreich Westfalen mit der Einführung der neuen Gemeindeverfassung die Aufhebung der rechtlichen Sonderstellung der Stadt und die Einführung der Gewerbefreiheit. Zur Entwicklung des rechtlichen Stadtbegriffs im 19. Jahrhundert vgl. allgemein: Matzerath, H., *Von der Stadt zur Gemeinde, Zur Entwicklung des rechtlichen Stadtbegriffs im 19. und 20. Jahrhundert*, in: Archiv für Kommunalwissenschaften, 13 (1974), S. 17–45; für Westfalen: Haase, C., *Die Entstehung der westfälischen Städte*, 2. Auflage, Münster 1965, S. 190 ff.

Der Einführung der Gewerbefreiheit kommt, dies haben die Untersuchungen von: Henning, F. W., *Die Einführung der Gewerbefreiheit und ihre Auswirkungen auf das Handwerk in Deutschland*, in: Abel, W. u. Mitarb. (Hrsg.), *Handwerksgeschichte*, S. 142–172 deutlich gezeigt, für die langfristige Entwicklung des Handwerks kaum eine Bedeutung zu. Es wird vielmehr deutlich, daß mit der Einführung der Gewerbefreiheit eine Entwicklung zur Ausweitung gewerblicher Produktionsstandorte auf dem Lande legalisiert wird, die bereits im 18. Jahrhundert sehr weit fortgeschritten war, wie dies u. a. in den Arbeiten von Schremmer, *Standortausweitung der Warenproduktion*, Kaufhold, *Umfang und Gliederung*, nachgewiesen wird. Im 18. Jahrhundert wurde verschiedentlich von staatlicher Seite, allerdings mit unterschiedlichem Erfolg, versucht, die Territorialisierung des Gewerbes zu verhindern. Zu den allerdings erfolglosen Bemühungen im Hochstift Münster, das städtische Gewerbe durch Ausschaltung der ländlichen Konkurrenz zu fördern, vgl. u. a.: Reekers, S., *Beiträge zur statistischen Darstellung der gewerblichen Wirtschaft Westfalens um 1800*. Teil 1: *Münster und Paderborn*, in: Westfälische Forschungen, 17 (1964), S. 83–176.

⁵ Schremmer, *Standortausweitung*, S. 5–6

Als Fragestellung der vorliegenden Untersuchung ergibt sich damit: Bilden die Städte gegenüber den Gemeinden des flachen Landes unter sozialem wie wirtschaftlichem Aspekt noch eine geschlossene Gruppe? Wo befinden sich die gewerblich geprägten Bereiche, wo solche, die von agrarischer Struktur bestimmt werden, an welchen Orten sind Dienstleistungen lokalisiert? Inwieweit interferieren die auf diese Weise gewonnenen Einheiten mit der klassischen Stadt-Land-Gliederung?

1.2 Methodisches Vorgehen

Der Gang der Untersuchung sei im folgenden kurz dargestellt. Zunächst wird der Untersuchungsraum und das zur Verfügung stehende Datenmaterial vorgestellt; die für diese Untersuchung ausgewählten Variablen zur Ausstattung im Bereich des Gewerbes und der Dienstleistungen werden beschrieben. Im nächsten Abschnitt soll nach einer einführenden Darstellung der Faktorenanalyse an Hand des für den Untersuchungsraum vorliegenden Variablensatzes geprüft werden, in welchem Umfang die Faktorenanalyse im Rahmen regionaler Strukturanalysen ein sinnvolles Verfahren darstellt. Aus der sich in diesem Zusammenhang ergebenden Erkenntnis in die begrenzten Einsatzmöglichkeiten des Verfahrens, wird in einem weiteren Abschnitt eine Strukturierung und Beschreibung des Datenmaterials über Klassifikationsverfahren versucht, dabei werden hierarchische und iterative Klassifikationsmethoden verwendet. Die hinsichtlich ihrer Struktur ähnlichen Gemeinden werden auf diese Weise zu Gruppen zusammengefaßt, über die Beschreibung dieser Gruppen werden Ausstattungsunterschiede deutlich gemacht.

2. Untersuchungsraum und sozial-ökonomische Merkmale

Die Analyse der sozial-ökonomischen Struktur städtischer wie ländlicher Gemeinden soll beispielhaft im östlichen und Kern-Münsterland durchgeführt werden. Das Untersuchungsgebiet besteht aus den alten preußischen Landkreisen Beckum, Lüdinghausen, Münster und Warendorf sowie der kreisfreien Stadt Münster; es ist etwa 2.920 km² groß und besitzt 1816 rund 136.000 Einwohner⁶. Von den 95 selbständigen politischen Gemeinden sind 15 als Städte im weiteren Sinne zu bezeichnen, wobei sich diese wiederum in 8 Städte und 7 Wigbolde/Flecken aufgliedern. In der Terminologie des frühen 19. Jahrhunderts spiegelt sich noch die spätmittelalterliche städtische Tradition wider⁷. Diese städtischen Siedlungen sind die Schwerpunkte

⁶ Kohl, W., *Politische Gliederung 1804, 1809/1811, 1818*, in: *Geschichtlicher Handatlas von Westfalen*, 1. Lieferung, Münster 1975.

⁷ Vgl. Stoob, H., *Minderstädte, Formen der Stadtentstehung im Spätmittelalter*, in: VSWG, 46 (1959), S. 1–28.

der Bevölkerung, in ihnen leben ungefähr 28 Prozent der Bevölkerung. Allerdings bestehen zwischen den Städten bedeutende Unterschiede hinsichtlich der Einwohnerzahl. Münster als ehemalige Hauptstadt des gleichnamigen geistlichen Territoriums ist die mit Abstand größte Stadt (15.088 Einwohner) und damit von der Einwohnerzahl her etwa viermal so groß wie die größte Landstadt des Untersuchungsgebietes, Warendorf, die immerhin 3614 Einwohner hat, und sich schon von der Größe her von einer aus vier Städten bestehenden Gruppe (Ahlen, Beckum, Telgte, Werne mit Einwohnerzahlen zwischen 1500 und 2200) abhebt. Die übrigen Städte haben deutlich weniger als 1500 Einwohner und gehören, abgesehen von Lüdinghausen und Sendenhorst, zum sogenannten *Minderstadt*-Typ. Die Untersuchungsgemeinden sind in *Tabelle 1* zusammengefaßt.

Tabelle 1:
Die Gemeinden des Untersuchungsraumes im östlichen Münsterland

Nr.	Gemeinde	Nr.	Gemeinde
1	Münster	48	Ennigerloh
2	St. Mauritz	49	Oelde, Std.
3	Lamberti	50	Oelde, Ksp.
4	Überwasser	51	Stromberg
5	Handorf	52	Wadersloh
6	Amelsbüren	53	Liesborn
7	Hiltrup	54	Diestedde
8	Telgte, Std.	55	Herzfeld
9	Telgte, Ksp.	56	Lippborg
10	Westbevern	57	Ahlen
11	Wolbeck, Std.	58	Altahlen
12	Wolbeck, Ksp.	59	Neuahlen
13	Angelmodde	60	Heessen
14	Alverskirchen	61	Dolberg
15	Albersloh	62	Lütke Untrop
16	Rinkerode	63	Beckum, Std.
17	Nottuln	64	Beckum, Ksp.
18	Schapdetten	65	Sünninghausen
19	Roxel	66	Vellem
20	Albachten	67	Vorhelm
21	Bösensell	68	Enniger
22	Nienberge	69	Sendenhorst, Std.
23	Havixbeck	70	Sendenhorst, Ksp.
24	Greven	71	Drensteinfurt, Std.
25	Gimbte	72	Drensteinfurt, Ksp.
26	Saarbeck	73	Herbern
27	Ladbergen	74	Bockum
28	Ostbevern	75	Hövel
29	Milte	76	Walstedde
30	Einen	77	Werne, Std.

31	Lienen	78	Werne, Ksp.
32	Warendorf	79	Stockum
33	Altwarendorf	80	Kapelle
34	Neuwarendorf	81	Südkirchen
35	Füchtorf	82	Bork
36	Sassenberg	83	Altlinen
37	Harsewinkel, Std.	84	Lüdinghausen, Std.
38	Harsewinkel, Ksp.	85	Lüdinghausen, Ksp.
39	Marienfeld	86	Senden
40	Greffen	87	Appelhülsen
41	Beelen	88	Venne/Ottmarsbocholt
42	Freckenhorst, Std.	89	Ascheberg
43	Freckenhorst, Ksp.	90	Nordkirchen
44	Everswinkel	91	Selm
45	Hoetmar	92	Olfen, Std.
46	Westkirchen	93	Olfen, Ksp.
47	Ostenfelde	94	Seppenrade

Zur Beschreibung der in diesen Gemeinden vorhandenen sozio-ökonomischen Ausstattung werden nun operationale Größen abgeleitet, welche Informationen über die gewerbliche Ausstattung, den Grad der Versorgung der Bevölkerung liefern. Solche Daten liegen auf Gemeindebasis in den amtlichen preußischen statistischen Erhebungen aus dem ersten Viertel des 19. Jahrhunderts vor. Quellen dieser Art eignen sich hervorragend für großräumige vergleichende Untersuchungen, weil sich die in ihnen enthaltenen Daten auf einen einheitlichen, zeitlich eng begrenzten Zeitraum beziehen und in aller Regel auch an allen Plätzen einer größeren Region nach nahezu gleichen Prinzipien erhoben worden sind.

Die amtliche Statistik der ersten Hälfte des 19. Jahrhunderts⁸, nach der Gründung des *Königlichen Preußischen Statistischen Büros* 1805, bringt gegenüber den zahlreichen privaten, halbamtlichen und amtlichen statistischen Erhebungen und Darstellungen des 18. Jahrhunderts einen wesentlichen Fortschritt, der vor allem in dem Bemühen um systematische Gliederung und um Vollständigkeit des Erhebungsmaterials sowohl in sachlicher als auch in räumlicher Hinsicht gegeben ist. Doch sollte einer Auswertung der frühen amtlichen statistischen Erhebungen in jedem Fall eine sorgfältige Quellenanalyse vorangehen.

Schwierigkeiten bei der Interpretation der frühen Statistik ergeben sich zum einen aus der häufig festzustellenden mangelnden Sorgfalt bzw. Willkür der erhebenden Behörden, zum anderen beruhen sie darauf, daß Begriffs- und Merkmalsdefinitionen der zeitgenössischen Statistik erheblich von den in der modernen Statistik

⁸ Einen Abriß der Geschichte der preußischen Statistik gibt: Büsch, O., *Industrialisierung und Gewerbe im Raum Berlin/Brandenburg 1800–1850, Eine empirische Untersuchung zur gewerblichen Wirtschaft einer hauptstadtgebundenen Wirtschaftsregion*, Berlin 1971, S. 152–204. Vgl.: Blenk, E. (Hrsg.), *Das Königliche Statistische Bureau im ersten Jahrhundert seines Bestehens 1805 bis 1905*, Erster Teil der Festschrift, Berlin 1905.

üblichen Begriffen abweichen. Insbesondere Zeitreihenanalysen werden im allgemeinen dadurch erschwert, daß Begriffe und Erhebungsschemata wenig konstant sind, immer wieder ist nämlich von seiten der Erhebungsbehörde versucht worden, durch Veränderung und Erweiterung der Begriffe, durch Einführung zusätzlicher Erhebungstatbestände die Statistik der sich schnell verändernden wirtschaftlichen Entwicklung anzupassen. Eine besondere Schwierigkeit ergibt sich ferner daraus, daß in der Regel nicht eindeutig feststellbar ist, ob es sich bei den jeweiligen statistischen Daten um gewerbe- oder berufsstatistische Angaben handelt. Trotz aller Schwierigkeiten und Bedenken sind die statistischen Quellen aus der ersten Hälfte des 19. Jahrhunderts wichtige Unterlagen zur Erforschung regionaler wie lokaler, sozialer und ökonomischer Strukturen.

Die *Statistische und Topographische Tabelle des Jahres 1816*⁹ gehört in die Reihe der bis 1858 alle drei Jahre erstellten preußischen Gewerbetabellen. Ihr Erhebungsschema basiert im wesentlichen auf den Vorschlägen von J.G. Hoffmann und ist in abgeänderter Form erstmals 1810 für eine statistische Erhebung verwendet worden. Immer wieder den sich verändernden wirtschaftlichen Gegebenheiten angepaßt, bildet es bis zur Zollvereins-Gewerbestatistik von 1861 die Grundlage sämtlicher berufs- und betriebsstatistischen Erhebungen.

Die *Statistische und Topographische Tabelle von 1816* gliedert sich in 434 Spalten, die in 5 Hauptabteilungen zusammengefaßt werden: *Gebäude, Bevölkerung, Unterrichts-Anstalten, Polizey-Anstalten, Erwerbsmittel*. Die Abteilung *Erwerbsmittel* gliedert sich wiederum in 273 Kategorien, die in 13, allerdings recht inhomogene Gruppen zusammengefaßt sind¹⁰. Bereits Zeitgenossen empfinden dieses Gliederungsschema als wenig befriedigend, so stellt z. B. Dieterici fest: „Die Kolonnen betreffen ... teils Handwerk, teils Fabrikation, teils Handel, teils Rohproduktion. Man hat die gewerbliche Tätigkeit der Nation nirgends beisammen, muß die zahlreichsten Handwerker ... und andere an den verschiedensten Stellen suchen, die Mühlenwerke sind voneinander getrennt, die Fabrikationen stehen nicht beieinander und ganz unbedeutende sind wie Hauptkategorien behandelt“¹¹. Es erscheint daher sinnvoll, vor einer Auswertung dieser Tabelle die 273 Kolonnen der *Hauptabteilung Erwerbsmittel* unter Einschluß einiger weiterer Daten aus anderen *Hauptabteilungen* in der Form umzugestalten, daß Berufe und Berufsgruppen nach wirtschaft-

⁹ Staatsarchiv Münster, Reg. Münster Nr. 1684.

¹⁰ Die Statistik von 1816 unterscheidet folgende Hauptgruppen: Viehzucht, Fleisch- u. Fettwaren, Bearbeitung von Fettwaren, Abdeckereien, Zubereitung u. Verarbeitung von Leder, Arbeiten in Holz, Getreide und Fabrikate daraus, Erwerb aus anderen Vegetabilien, Zubereitung u. Benutzung der Mineralien, Weberei, zugehörige Vorbereitungen u. Appreturen auch verwandte Gewerbe, verschiedene andere Gewerbe (Arbeiten in Papiermasse, Siegellack-Bleistift-Tonstein- und Schreibfederfabriken, Arbeiten in Federn, Haar, Horn, Schellack u. Kork auch Putzwaren, Mechanische Künste), Erwerbsmittel in Bezug auf Wissenschaften, schöne Künste u. öffentliches Vergnügen, Erwerbsmittel in Bezug auf Handel u. Verkehr im Allgemeinen.

¹¹ Dieterici, C.F.W. (Hrsg.), *Tabellen und amtliche Nachrichten über den preußischen Staat für das Jahr 1846*, Bde. 1–6, Berlin 1852–1855, hier Bd. 6 b, S. 1000, zit. nach: Büsch, *Industrialisierung und Gewerbe*, S. 153.

lichen sowie sozialen Schichtungsmerkmalen einander zugeordnet werden. Die Berufsklassifizierung des *Statistischen Bundesamtes* von 1961/1970¹² und das von O. Büsch für die statistische Erhebung von 1849 entwickelte Gliederungsschema¹³ stellen dafür die Grundlage zu dieser Neuordnung. Zwei Berufsabschnitte sollen beispielhaft für viele andere in der nachstehenden *Tabelle 2* in Auszügen wiedergegeben werden.

Tabelle 2
Neuordnung der Statistischen und Topographischen Tabelle von 1816:

<i>Berufsklassifikation des Statistischen Bundesamtes 1970</i>	<i>Schema der Statistischen und Topographischen Tabelle des Jahres 1816 Gewerbebezeichnung</i>	<i>Spalte</i>
<i>Berufsabschnitt III f</i>		
Metallerzeuger, -	Eisenhüttenwerke und dahin gehörige Anlagen, als:	
Metallbearbeiter	hohe Öfen, Frischfeuer, Stab-, Reck- und Zainhämmer,	
	Blech- und Drahtwerke	251
	Zahl der dabei beschäftigten Arbeiter	252
	Kupfer- und Messinghüttenwerke und dahin gehörige	
	Anlagen, als: Rohkupferhütten, Kupfersaigerhütten,	
	Kupferhämmer, . . .	253
	Zahl der dabei beschäftigten Arbeiter	254
	Blei-, Zinn-, Galmei-, Zink- und Arsenhüttenwerke	
	(Schrootgießereien und Bleiplattenfabriken)	255
	Zahl der dabei beschäftigten Arbeiter	256
	Stahl- u. Eisenwaren-Fabriken (einschließlich Ge-	
	wehrfabriken)	257
	Zahl der dabei beschäftigten Arbeiter	258
	Schriftgießereien, Zahl derselben	377
	Zahl der dabei beschäftigten Arbeiter	378
	Zinngießerei und Zinnknopfmacher, Meister	275
	Gehülfen und Lehrlinge	276
	Gelb- und Rothgießer, Glockengießer, Stückgießer,	
	Meister	265
	Gehülfen und Lehrlinge	266

¹² Statistisches Bundesamt (Hrsg.), 1961, 1970.

¹³ Vgl.: Büsch, *Industrialisierung und Gewerbe*, S. 159 ff.; auch; Kaufhold, K.H., *Inhalt und Probleme einer preußischen Gewerbestatistik vor 1860*, in: Bog, I., Franz, G. u. a. (Hrsg.), *Wirtschaftliche und soziale Strukturen im säkularen Wandel*. Festschrift für W. Abel zum 70. Geburtstag, Hannover 1974, Bd. 3, S. 707–719.

Fortsetzung, Tab. 2

<i>Berufsabschnitt III 1</i>		
Lederherstellung,	Lederbearbeitung aller Art, als Roth-, Weiß- und	
Leder- und Fellverar-	Sämisich-Gerberei, Korduan-, Saffiangefärbte	
beitung	Leder- und Pergament-Bereitung, Lederthauer,	
	Loh- und Leder-Walkmühlen, Zahl der Meister	
	oder Fabrikinhaber	192
	Zahl der Gehülfen, Lehrlinge und Fabrikarbeiter	193
	Handschuhmacher und Beutler, Meister	194
	Gehülfen und Lehrlinge	195
	Rierner und Sattler, Meister	196
	Gehülfen und Lehrlinge	197
	Schuhmacher, Pantoffelmacher und Schuhflicker,	
	Meister	198
	Gehülfen und Lehrlinge	199
	Kürschner und Rauchwaarenhändler, auch Zobel-	
	färber, Meister und Herren	200
	Gehülfen und Lehrlinge	201

Bevor der in der Analyse eingehende Merkmalskanon vorgestellt wird, sollten noch einige Aspekte zur Struktur der Merkmale dargelegt werden.

Ziel der Untersuchung ist bekanntlich die Beschreibung der in den Gemeinden im gewerblich-handwerklichen Bereich und im Bereich der Dienstleistungen vorhandenen Ausstattung; es ist darum nun die Aufgabe, solche Merkmale abzuleiten, die Rückschlüsse auf den vorhandenen Bestand an Infrastruktureinrichtungen wie auch den Grad der Versorgung der Bevölkerung gestatten. Die *Verhältniszahl* stellt eine solche Größe dar. Mit ihrer Hilfe ist es möglich, den vorhandenen Versorgungsgrad eines bestimmten Bedarfsträgers mit spezifischen Einrichtungen anzugeben. Wenn Verhältniszahlen Verwendung finden, dann 1. mit dem Ziel, durch den Bezug auf potentielle Nutzer das Niveau in der Versorgung und Ausstattung besser erfassen zu können, und 2., um auf diese Weise den Einfluß der Siedlungsgröße zu verringern bzw. ganz ausschalten zu können. Es ist nämlich davon auszugehen, daß in der Regel jedes Merkmal in irgendeiner Weise von der Ortsgröße abhängt; eine Verwendung von Absolutdaten würde somit zu trivialen Ergebnissen führen¹⁴.

Die *Ortsgröße*, gemessen in der Einwohnerzahl, ist als ein wesentliches, den Bedarf bestimmendes Element zu betrachten. In der überwiegenden Mehrzahl der Merkmale ist dieser Aspekt durch Relativierung der jeweiligen Variablen auf die Einwohnerzahl berücksichtigt worden. Bei einigen Merkmalen wird jedoch durch die Einwohnerbezogenheit der Daten ein *schiefes* Bild in der Hinsicht entstehen, als sich in den so gewonnenen Werten eine Überversorgung in bestimmten Bereichen auszudrücken scheint. Tatsächlich drückt sich jedoch in diesen Werten nicht nur der

¹⁴ Vgl. hierzu etwa: Forst, H.T., *Zur Klassifizierung von Städten nach wirtschafts- und sozial-statistischen Strukturmerkmalen*, Würzburg 1974, S. 13–14.

Versorgungsgrad der ortsansässigen Bevölkerung aus, sondern auch die Verflechtung des Ortes mit seinem Umland. Obwohl bei diesen *zentralen* Merkmalen ein Bezug auf die Ortsbevölkerung nicht unproblematisch ist, wird auch in diesen Fällen keine andere Relativierung vorgenommen, weil sich keine *regionalen* Bezugswerte ermitteln lassen.

In der *Tabelle 3* sind die für die Problemstellung dieser Untersuchung wichtigen Merkmale zusammengestellt. Sie sind aus der oben beschriebenen *Topographischen und Statistischen Tabelle* entweder direkt entnommen worden oder durch Zusammenfassung von verschiedenen Merkmalen entstanden.

Tabelle 3
Merkmale zur sozial-ökonomischen Struktur

<i>Var.Nr.</i>	<i>1. Variablen zur Bevölkerungsstruktur</i>
1	Bevölkerungsdichte 1816
2	Bevölkerungsverdichtung 1795–1843
3	Protestanten je 1000 Einwohner
4	Juden je 1000 Einwohner
5	unter 8jährige je 1000 Einwohner
6	8–14jährige je 1000 Einwohner
7	15–60jährige je 1000 Einwohner
	<i>2. Variablen zur Flächennutzung</i>
8	Ackerland, Wiese/Weide in % an der Gemeindefläche
9	Gartenland in % an der Gemeindefläche
	<i>3. Variablen zur Erwerbsstruktur</i>
	<i>3.1 Landwirtschaft</i>
10	landwirtschaftliche Betriebe je Einwohner
11	landwirtschaftliche Betriebe 15 Morgen je landwirtschaftlicher Betrieb gesamt
	<i>3.2 produzierendes Gewerbe</i>
	<i>Steinbearbeitung, Glas, Keramik</i>
12	Steinmetzarbeit, Steingut ¹⁵
13	Porzellan, Glas: Beschäftigte je Einwohner
	<i>Bauhaupt- und Baunebengewerbe</i>
14	Maurerei, Zimmerei, Dachdeckerei, Rietbinderei: Beschäftigte je Einwohner
15	Glaseri, Malerei
16	Stukkateur
	<i>Metallverarbeitendes Gewerbe</i>
17	Grobschmiede
18	Kupferschmiede
19	Metallverformung, Feinblechbearbeitung

¹⁵ Alle Angaben zur Erwerbsstruktur beziehen sich, falls nicht anders angegeben, auf die jeweilige Wohnbevölkerung.

20	Feinmechanik, Uhren, Schmuck
21	Kammsetzer, Streichenmacher
22	Metallguß
23	<i>Papierverarbeitung, Druckereigewerbe</i>
24	<i>Farbstoff- und Seifenherstellung</i>
	<i>Holzverarbeitendes Gewerbe</i>
25	Holzschuhherstellung
26	Tischlerei
27	Böttcherei, Korbwarenherstellung
28	Kunstdrechserei
	<i>Textil- und Bekleidungsgewerbe</i>
29	Lederverarbeitung
30	Schuhmacherei
31	Kürschnerei
32	Schneiderei
33	Hut- und Putzmacherei, Posamentiererei, Raumausstattung
34	Seilerei
35	Spinnerei: Spindeln auf Wolle und Baumwolle je Einwohner
36	Weberei: Webstühle auf Wolle, Baumwolle, Herrenhuter Zeug je Einwohner
37	Weberei: Webstühle auf Leinen je Einwohner
38	Weberei: Webstühle in Nebenbeschäftigung je Einwohner
39	Textilzubereitung
	<i>Nahrungs- und Genußmittelgewerbe</i>
40	Bäckerei, Metzgerei
41	Brennerei, Brauerei, Stärkeherstellung
42	Tabakwarenherstellung
	<i>3.3 Handel und Dienstleistungen</i>
43	Großhandel
44	Handel mit Material-, Gewürz-, Ellenwaren
45	Handel mit Metallwaren
46	Handel mit Getreide, Wein
47	Buchhandel
48	Bilder-, Devotionalienhandel
	<i>Ambulanter Handel</i>
49	Garnhandel
50	Höker
51	Verkehr
52	Körperpflege
	<i>Gaststätten</i>
53	Gaststätte für Personen aus gebildeten Ständen je Einwohner
54	Gasthöfe in der Stadt je Einwohner
55	Krüge auf dem Lande je Einwohner

3.1 Faktorenanalyse und sozial-ökonomische Struktur

Im folgenden Abschnitt soll nun die Frage geprüft werden, in welchem Umfang die Faktorenanalyse sinnvoll bei der Untersuchung der Gewerbe- und Dienstleistungsausstattung der Gemeinden verwendet werden kann. Als multivariates statistisches Verfahren ist die Faktorenanalyse darauf gerichtet, zwischen den Variablen bestehende Zusammenhänge aufzudecken.

Als Grundlage des faktorenanalytischen Vorgehens¹⁶ ist eine $n \times m$ Datenmatrix X gegeben, die für die n Untersuchungsobjekte jeweils die Ausprägungen der m Merkmale enthält.

$$X = \begin{bmatrix} x_{11} & \dots & x_{1j} & \dots & x_{1m} \\ x_{i1} & \dots & x_{ij} & \dots & x_{im} \\ x_{n1} & \dots & x_{nj} & \dots & x_{nm} \end{bmatrix}$$

Jede Zeile $x_i = (x_{1i} \dots x_{im})$ $i = 1, \dots, n$ beschreibt die Eigenschaften der Objekte O_i , die Spalten X^j

$$X^j = \begin{bmatrix} x_{1j} \\ \vdots \\ x_{nj} \end{bmatrix} \quad j = 1, \dots, m$$

stellen die Ausprägungen der Variablen x_j dar.

Nun ist aber, da die Variablen ein gemeinsames Objekt (im vorliegenden Fall: Gemeinden in einer fixierten Region) beschreiben, davon auszugehen, daß diese Variablen nicht völlig unabhängig voneinander variieren, sondern daß einzelne Variablen oder Variablengruppen vorhanden sind, die stärker miteinander korrelieren, also ähnliche oder gemeinsame Informationen enthalten.

Von diesen Korrelationen zwischen den Variablen geht die Faktorenanalyse aus. Es wird angenommen, daß es, im allgemeinen unabhängige, direkt nicht meßbare Größen (Variablen höherer Ordnung) gibt, die in irgendeiner Weise hinter den beobachteten Merkmalen stehen bzw. sie bewirken.

¹⁶ Auch bei der Lösung eines Klassifikationsproblems ist im allgemeinen eine Datenmatrix X gegeben.

Die Faktorenanalyse dient 1. der Bestimmung einer möglichst kleinen Zahl Faktoren, 2. der Bestimmung eines Zusammenhangsmaßes zwischen Variablen und Faktoren, 3. der Bestimmung von Werten, die die Objekte auf den Faktoren einnehmen¹⁷.

Im folgenden ist daher zu untersuchen, in welchem Umfang es mit Hilfe der Faktorenanalyse gelingt, den umfangreichen und unübersichtlichen Datensatz zur Ausstattung im Bereich der Infrastruktur des Gewerbes und der Dienstleistungen zu beschreiben, zu strukturieren sowie schließlich, bei Tolerierung eines gewissen Informationsverlustes, auf wenige, interpretierbare Dimensionen (*Faktoren*) zu reduzieren.

In seinem zusammenfassenden Überblick über die Einsatzmöglichkeiten der Faktorenanalyse in der Untersuchung regionaler Strukturen unterscheidet Klemmer¹⁸ zwei Aufgabenfelder: 1. die Faktorenanalyse als ein Verfahren zur Datenreduktion und 2. die Faktorenanalyse als eine Technik zur Formulierung komplexer Indikatoren. Auf das erste Aufgabenfeld wird noch im Zusammenhang mit der Darstellung der Clusteranalyse näher eingegangen werden, an dieser Stelle interessiert besonders die zweite Verwendungsmöglichkeit der Faktorenanalyse; dabei stehen die folgenden Probleme im Vordergrund des Interesses:

1. Die Untersuchung der Frage, in welchem Umfang das Datenmaterial sich auf eine Dimension reduzieren läßt¹⁹, d. h., es ist zu prüfen, ob sich ein Hauptfaktor extrahieren läßt, der einen großen Teil der Varianz erklärt. Für den Fall, daß alle Indikatoren einer Dimension zugehören, können die *Faktorenladungen* „als Gewichte im Sinne der Indexkonstruktion“²⁰ verwendet werden. Auf diese Weise wird über die Faktorenanalyse eine homogene Skala aus verschiedenen Indikatoren erzeugt. Im Sinne dieser Untersuchung ergäbe sich eine Skala der Infrastrukturausstattung im Bereich von Gewerbe und Dienstleistung, nach der die Gemeinden zu gruppieren sind.

2. Für den Fall, daß das Datenmaterial nicht eindimensional ist, ist nach einer sinnvollen Interpretation der den Datensatz beschreibenden Faktoren (*Dimensionen*) zu suchen. In einem solchen Fall kommt der Faktorenanalyse eine Hilfsfunktion zu, indem sie das Datenmaterial für eine nachfolgende Analyse, in dieser Untersuchung etwa ein Verfahren zur Ermittlung von Gemeindetypen, aufbereitet. Diese zuletzt genannte Analysemöglichkeit soll in einem späteren Abschnitt dieser Arbeit darge-

¹⁷ Vgl.: Clauß, G. und Ebner, H., *Grundlagen der Statistik für Psychologen, Pädagogen und Soziologen*, 2. neu bearb. u. erw. Auflage, Zürich, Frankfurt 1975, S. 353 ff. Zur Faktorenanalyse siehe: Überla, K., *Faktorenanalyse, Eine systematische Einführung für Psychologen, Mediziner, Wirtschafts- und Sozialwissenschaftler*, 2. Aufl., Berlin, Heidelberg, New York 1971; oder: Weber, E., *Einführung in die Faktorenanalyse*, Stuttgart 1974.

¹⁸ Klemmer, P., *Die Faktorenanalyse als Instrument der empirischen Strukturforchung*, in: *Methoden der empirischen Regionalforschung* (1. Teil), Hannover 1973, S. 131–146, hier S. 143–146.

¹⁹ Vgl. hierzu allgemein: Werner, R., *Soziale Indikatoren und politische Planung, Einführung in Anwendungen der Makrosoziologie*, Reinbek 1975, hier bes. S. 199 ff.

²⁰ Werner, *Soziale Indikatoren*, S. 199.

stellt werden²¹. Zunächst geht es um den Einsatz der Faktorenanalyse auf der Basis einer 94 x 55 Datenmatrix nach z-Transformation.

Von diesem Datensatz wird zunächst die Korrelationsmatrix (*Produkt-Moment-Korrelation*) ermittelt; in dieser Matrix werden die $m(m-1)/2$ möglichen Beziehungen zwischen den 55 Merkmalen zusammengefaßt²². Da die Korrelationsmatrix alle für die Faktorenextraktion notwendigen Informationen enthält, ergeben sich über eine Interpretation der Korrelationsmatrix erste Anhaltspunkte für Zusammenhänge innerhalb des Datensatzes. Von den 1485 Korrelationskoeffizienten zwischen den Variablen sind 557 (= 37,5 %) signifikant von Null verschieden²³. In einer genaueren Betrachtung der Korrelationsmatrix zeigt sich, daß neben hohen Koeffizienten auch zahlreiche niedrige Werte auftreten. 29 Korrelationskoeffizienten übersteigen den Wert von /0.8/.

Insgesamt sind 157 Koeffizienten größer als /0.5/. Bei gleichzeitiger Beachtung der Merkmale mit nur geringen wechselseitigen Korrelationen ergeben sich in der Matrix deutliche Variablengruppierungen: Es ergibt sich eine Merkmals-Gruppe, die die gewerbliche Grundausrüstung der Gemeinden beschreibt; diese zeichnen sich dadurch aus, daß sie untereinander, wie auch mit anderen Merkmalen durchweg nur sehr gering korrelieren. Die andere Variablengruppierung, erkennbar an der erhöhten wechselseitigen Korrelation, gliedert sich in zwei Kategorien: Während die eine Kategorie Spezialgewerbe der Textilbranche umfaßt, besteht die andere aus Variablen des hochspezialisierten Gewerbebereiches. Es ist zu erwarten, daß diese skizzierten Zusammenhänge in einer Faktorenanalyse bestimmend für die Faktoren sein werden.

Die folgenden Untersuchungen werden nur mit dem einfachen Modell der *Hauptkomponentenanalyse* durchgeführt, dabei wird im Gegensatz zur Faktorenanalyse im engeren Sinne auf die Schätzung der Kommunalitäten verzichtet.

Mit der *Hauptachsenmethode* werden 14 Faktoren extrahiert, deren Eigenwert > 1.0 ist. Die Ergebnisse sind in *Tabelle 4* zusammengefaßt; der erste Faktor erklärt 29,2 % der Gesamtvarianz, der zweite nur noch 8,8 %. Die 14 Faktoren stellen insgesamt nur 82,5 % der Gesamtvarianz dar. Im Merkmalssatz ist somit ein bestimmender Faktor nicht vorhanden. Damit erscheint es nicht mehr sinnvoll, die Variablen zur sozio-ökonomischen Struktur der Städte zu einer Kennziffer zusammenzufassen.

²¹ Vgl. Abschnitt. 3.2.

²² Sämtliche Rechenarbeiten sind im Rechenzentrum der Universität Münster auf einer IBM 360/50 durchgeführt worden.

Neben dem Programm *FACTOR* des Programmpaketes *SPSS* sind Programme zur Faktorenanalyse und zur Clusteranalyse verwendet worden, die von Herrn Dr. D. Steinhausen und Herrn Ch. Zeidler entwickelt worden sind. Ihnen sei an dieser Stelle herzlich gedankt.

²³ Der Signifikanztest wird mit Hilfe der Testgröße $F = \frac{r^2(n-2)}{1-r^2}$ vorgenommen. Für $n = 94$

und $\alpha = 0,05$ werden alle Korrelationskoeffizienten größer /0,203/ als signifikant angesehen. Vgl. auch Sachs, L., *Statistische Auswertungsmethoden*, 3. neu bearb. u. erweiterte Auflage, Berlin, New York, Heidelberg 1971, S. 329 ff.

Tabelle 4

Faktorextraktion nach dem Hauptkomponentenmodell (Hauptachsenmethode)

Faktor	Eigenwert	Erklärte Varianz in % je Faktor	kumuliert
1	16,1	29,2	29,2
2	4,9	8,8	38,0
3	4,6	8,3	46,3
4	3,2	5,8	52,1
5	2,6	4,8	56,9
6	2,3	4,3	61,2
7	2,1	3,9	65,1
8	1,9	3,5	68,6
9	1,6	2,9	71,5
10	1,5	2,7	74,2
11	1,3	2,4	76,6
12	1,2	2,1	78,8
13	1,1	2,0	80,7
14	1,0	1,9	82,6

Es ist darum nun zu untersuchen, inwieweit sich die Dimensionen des Variablenkanons auf wenige interpretierbare Faktoren sinnvoll reduzieren lassen. Dazu werden nach Bestimmung der Zahl der signifikanten Faktoren diese orthogonal nach dem Varimax-Kriterium auf Einfachstruktur rotiert.

Nach dem *Scree-Test* ergeben sich 6 signifikante Faktoren, die 61,2 % der Gesamtvarianz erklären. Bevor das rotierte Faktorenmuster interpretiert werden kann, ist es erforderlich, dieses nach dem *Bargmann-Test*²⁴ auf Einfachstruktur zu überprüfen. Dazu wird so vorgegangen, daß für jeden Faktor j die Zahl der Ladungen ermittelt wird, für die gilt

$$|a_{ij}| > 0,10 \cdot h_i; \text{ wobei } a_{ij} = \text{Ladung der Variablen } i \text{ auf den Faktor } j \\ h_i = \text{Kommunalität der Variablen } i$$

Die Analyse der Matrix der rotierten Faktorladungen ergibt dabei folgendes Bild:

Faktor	1	2	3	4	5	6	Σ
Anzahl der Faktorladungen	9	16	27	16	11	27	106

Von den 350 Faktorenladungen liegen 106 (= 32,1 %) innerhalb des Bereiches von $0,10 \cdot h_i$ um die Hyperebenen. Bei 6 Faktoren und 55 Variablen beträgt für $d = 0,05$ die Zahl der Ladungen je Faktor 19, für die $a_{ij} < 0,10 \cdot h_i$ ²⁵ gilt.

In der vorliegenden Matrix besitzen demnach nur die Faktoren 4 und 6²⁶ Ein-

²⁴ Überla, *Faktoranalyse*, S. 184 ff.

²⁵ Vgl. Überla, *Faktorenanalyse*, 1971, S. 374.

²⁶ Für $\alpha = 0,25$ liegt die kritische Anzahl von Nulladungen bei 16; bei diesem Signifikanzniveau besitzen die Faktoren 2, 3, 4 und 6 Einfachstruktur.

fachstruktur, d. h., nur diese Faktoren können als durch ausreichend viele Merkmale definiert angesehen werden und sind damit interpretierbar.

Der negative Bargmann-Test schließt eine Interpretation des Faktorenmusters eigentlich aus, unter Berücksichtigung der Ergebnisse des Bargmann-Tests für $\alpha = 0,25$ erscheint es jedoch gerechtfertigt, im vorliegenden Fall einige Hinweise auf die inhaltliche Bedeutung der Faktoren zu geben. Es sei aber auch darauf hingewiesen, daß bei insgesamt doch negativem Bargmann-Test die Interpretation der Faktoren in dem Sinne nicht möglich ist, daß sie als neue Variablen in weiteren Untersuchungen Verwendung finden.

Auf dem *ersten* Faktor laden Variablen aus dem Bereich des Handels mit Gütern des gehobenen Bedarfs (*Var. 22, 45, 46*), des Verkehrswesens (*Var. 51*). Weiter wird der Faktor bestimmt durch Variablen aus dem Bereich des Textilgewerbes (*Var. 31, 35, 36*), der Nahrungs- und Genußmittelproduktion (*Var. 23, 42*), des Gaststättengewerbes (*Var. 54*). Der *zweite* Faktor wird bestimmt durch die Bevölkerungsdichte und die Zahl der Buchhandlungen. Der *dritte* Faktor beschreibt die gewerbliche Grundausstattung. Der *vierte* Faktor wird bestimmt durch die Zahl der Beschäftigten in der Porzellanfabrikation und Glasbearbeitung sowie durch die Zahl der Beschäftigten im Devotionalienhandel. Der *fünfte* Faktor wird geprägt durch Variablen aus dem Bereich des relativ hoch spezialisierten Handwerks (*Var. 15, 18, 21*), der Genußmittelproduktion (*Var. 41*), des Handels (*Var. 44*) und Gaststättengewerbes (*Var. 53*). Dieser Faktor wird ferner bestimmt durch die hohe Ladung der *Var. 4* (Anteil Juden an der Wohnbevölkerung). Der *sechste* Faktor wird charakterisiert durch Variablen zur Altersstruktur (*Var. 6, 7*), sowie durch Variablen zur Größenordnung landwirtschaftlicher Betriebe (*Var. 11*).

Die Kurzbeschreibung der Faktoren läßt deutlich erkennen, daß die Faktoren nicht oder nur sehr schwer in einer Form zu interpretieren sind, daß sie als neue Variablen die Grundlagen für weitere Analysen abgeben können.

Zum Vergleich mit den Ergebnissen der Hauptkomponentenanalyse wird nun noch eine Faktorenanalyse im engeren Sinne, d. h. mit vorheriger Schätzung der Kommunalitäten durchgeführt; dabei wird das Faktorenmuster ebenfalls wiederum nach dem Varimax-Kriterium auf Einfachstruktur rotiert.

Zur Schätzung der *Kommunalitäten* werden in der Literatur verschiedene Verfahren angeführt²⁷. Da das Verfahren, die Kommunalitäten über das Quadrat des multiplen Korrelationskoeffizienten zu schätzen, an der Nicht-Invertierbarkeit der vorliegenden Korrelationsmatrix scheitert, wird bei der Kommunalitätenschätzung so vorgegangen, daß zunächst über eine Hauptkomponentenanalyse die Kommunalitäten bestimmt werden und die so ermittelten Werte dann als neue Schätzwerte in die Faktorenanalyse eingegeben werden.

Die auf diese Weise gewonnene Matrix der rotierten Faktorladungen zeigt weitgehend das gleiche Bild wie die nach dem Hauptkomponentenmodell gewonnene

²⁷ Vgl. Überla, *Faktorenanalyse*, 1971, S. 155 ff.; vgl. auch Holm, K., *Die Faktorenanalyse*, München 1976, S. 24, S. 71 ff., in: Holm, K. (Hrsg.), *Die Befragung*, 5 Bde. München 1976, hier Bd. 3.

Matrix. Die gleichen Variablen, die auch bei dem Verfahren ohne Kommunalitätenschätzung hohe Ladungen besitzen, haben auch in diesem Fall wiederum hohe Ladungen, allerdings in der Regel etwas niedriger. Der Bargmann-Test auf Einfachstruktur ist, wie auch bei der Hauptkomponentenanalyse, negativ.

Die Auszählung der Koordinatenhyperachse ergibt:

Faktor	1	2	3	4	5	6	Σ
Anzahl	10	10	13	23	27	22	105

Der negative Bargmann-Test in beiden Fällen läßt vermuten, daß die orthogonale Rotation dem vorliegenden Datensatz nicht angemessen ist und zugunsten einer schiefwinkligen Rotation aufzugeben ist. Nach dem *Oblimin-Kriterium*²⁸ sind verschiedene schiefwinklige Rotationen mit dem Ziel der Einfachstruktur durchgeführt worden. Die Analyse der Rotationen unter verschiedenen Rotationswinkeln (von fast orthogonalen bis zum extrem schiefwinkligen Fall) ergibt in jedem Fall einen negativen Bargmann-Test. Bemerkenswert ist in diesem Zusammenhang die Tatsache, daß das Muster der Faktoren sich in den verschiedenen schiefwinkligen Rotationen nur unwesentlich von dem bei orthogonaler Rotation unterscheidet. Aufgrund dieser weitgehenden Identität der Ergebnisse der verschiedenen Rotationsversuche ist der Schluß erlaubt, daß in dem vorliegenden Variablenatz keine dominierenden Dimensionen vorhanden sind.

Die Ergebnisse der bisherigen faktorenanalytischen Untersuchungen lassen sich in folgender Weise zusammenfassen:

1. Sowohl im einfachen Hauptkomponentenmodell als auch im Modell mit Kommunalitätenschätzung stellt sich heraus, daß kein Hauptfaktor zur sozio-ökonomischen Struktur zu extrahieren ist; mithin ist davon auszugehen, daß das Datenmaterial nicht eindimensional ist und somit auch nicht für einen homogenen Index verwendet werden kann.
2. Weder über die orthogonale noch über die schiefwinklige Rotation können sinnvoll zu interpretierende Faktoren ermittelt werden (keine Einfachstruktur). Die extrahierten Faktoren bilden damit keine ausreichende Grundlage für weiterführende Untersuchungen.

3.2.1 Klassifikationsverfahren

Die vorausgehenden, faktorenanalytisch ausgerichteten Untersuchungen haben gezeigt, daß auf diesem Wege keine Komprimierung der Informationen zu erreichen ist. Aus diesem Grunde soll nun im folgenden ein Weg versucht werden, der darauf abzielt, die Gemeinden des Untersuchungsraumes nach ihrer durch Merkmale zur sozialen wie ökonomischen Situation erfaßten Ähnlichkeit in homogene Gruppen

²⁸ Vgl. Nie, N.H., Hull, C.H., Jenkins, J.G., Steinbrenner, K., Bent, D.H.: *Statistical Package for the Social Sciences*, 2. Aufl., New York 1975, S. 485 ff.

aufzugliedern. Die Bildung homogener Gruppen aus einander ähnlichen Objekten bringt durch Beschreibung der sich ergebenden Gruppen die Möglichkeit, „die Struktur der betrachteten Objektmenge vereinfacht darzustellen und die Vielfalt der beobachteten Erscheinungsformen auf ein erträgliches, überschaubares Maß zu reduzieren. Das Prinzip der Klassenbildung erweist sich somit als eine Methode der Datenreduktion und insofern . . . als ein nützliches Hilfsmittel zur Erkenntnis neuer und unbekannter Zusammenhänge“²⁹.

Die Klassifikation, d. h. der Prozeß der Bildung von Klassen oder Gruppen, erfolgt unter der Verwendung mathematisch-statistischer Methoden, der sog. *Klassifikationsverfahren* oder *Cluster-Analysen*.

Grundlage sämtlicher clusteranalytischer Vorgehensweisen ist das Konzept des sog. *Merkmalsraumes*³⁰. Dabei wird unter Merkmalsraum im allgemeinen der Raum verstanden, der durch die für die Problemstellung zur Objektbeschreibung relevanten Merkmale aufgespannt wird. Einen solchen Raum darf man sich „in der Regel nicht als einen euklidischen Raum vorstellen und die üblichen Raumvorstellungen nicht ohne weiteres darauf übertragen“³¹.

Von der Problemstellung leitet sich jedoch nicht nur die Definition des Merkmalsraumes ab, sondern auch seine Struktur. Die Art, wie diese Struktur des Merkmalsraumes fixiert wird, d. h., welche Bedeutung den verschiedenen Merkmalskombinationen zugemessen wird, entscheidet darüber, ob sog. *künstliche* oder *natürliche* Klassen³² gebildet werden.

Im weiteren Verlauf der Untersuchung soll ausschließlich auf solche Verfahren, die mit der letzteren Gruppe, also den sog. natürlichen Klassen, zusammenhängen, Bezug genommen werden. Nach Sodeur werden „natürliche Typen . . . nach Gesichtspunkten gebildet, bei denen die Verteilung der Elemente im Merkmalsraum neben der theoretischen Relevanz der Merkmalskombinationen berücksichtigt wird oder sogar allein ausschlaggebend ist“³³. Die Bildung von sog. natürlichen Klassen oder Typen ist nur dann sinnvoll, wenn angenommen werden kann, daß die zu klassifizierenden Objekte eine ungleiche multivariate Verteilung im Merkmalsraum haben. Die Suche nach einer natürlichen Klassifikation ist ohne theoretische oder pragmatische Begründung solcher Diskontinuitäten im Verteilungsmuster der Objekte nicht zu rechtfertigen³⁴. Wie Schlosser gezeigt hat, sind faktorenanalytische Verfahren in diesem Zusammenhang wenig geeignet³⁵.

²⁹ Bock, H.H., *Automatische Klassifikation, Theoretische und praktische Methoden zur Gruppierung und Strukturierung von Daten (Cluster-Analyse)*, Göttingen 1974, S. 14.

³⁰ Zum Konzept des Merkmalsraumes vgl.: Sodeur, W., *Empirische Verfahren zur Klassifikation*, Stuttgart 1974, S. 11 ff. Die Clusteranalyse verwendet also ebenso wie das zuvor verwendete Verfahren der Faktorenanalyse die $n \times m$ Datenmatrix X , die für n Untersuchungseinheiten jeweils die Ausprägungen von m Variablen enthält.

³¹ Sodeur, *Empirische Verfahren*, S. 13.

³² Sodeur, *Empirische Verfahren*, S. 14–15.

³³ Sodeur, *Empirische Verfahren*, S. 15.

³⁴ Sodeur, *Empirische Verfahren*, S. 32.

³⁵ Die korrelations- bzw. faktorenanalytischen Konzepte sind auf die ausschließliche Erfas-

Bevor nun der vorliegende Datensatz mit Hilfe der Clusteranalyse untersucht werden soll, sind noch einige Bemerkungen zum Verfahren selbst zu machen. Klassifikationen können *monothetischer* oder *polythetischer* Art sein. Aufgrund einer monothetischen Klassifikation sich ergebende Klassen sind dadurch gekennzeichnet, daß „sämtliche zugehörigen Elemente gemeinsame Merkmalswerte tragen“³⁶, d. h. jede Klasse ist durch eine bestimmte Merkmalskombination (Merkmalsausprägung) charakterisiert. Das Konzept polythetischer Klassifikation zielt im Gegensatz zum monothetischen darauf ab, daß „die Elemente gleichen Typs eine wesentliche Anzahl gemeinsamer Merkmale besitzen . . . (aber darauf verzichtet wird), in welchen Merkmalen Übereinstimmung bestehen soll“³⁷. Dieses polythetische Konzept ist für die vorliegende Arbeit von besonderem Interesse, auf die weitere Darstellung des monothetischen Ansatzes soll deshalb an dieser Stelle verzichtet werden.

Innerhalb der polythetischen Gruppierungsarten stellen *hierarchisch-agglomerativen* Verfahren die bekanntesten Vorgehensweisen dar. Ihre Anwendung empfiehlt sich immer dann, wenn die Annahme *disjunkter* Klassen von der behandelten Fragestellung her vernünftig ist³⁸. Da eine solche Voraussetzung bei der hier behandelten Problematik sinnvoll ist, sollen hierarchisch-agglomerative Gruppierungsverfahren in Verbindung mit nicht-hierarchischen, aber ebenfalls disjunkten Methoden zur Anwendung kommen.

Der Ausgangspunkt hierarchisch-agglomerativer Verfahren ist eine Konfiguration von m Klassen zu je einem Objekt ($m = n$, wobei n = Anzahl der Beobachtungseinheiten). In den folgenden Fusionschritten werden dann – abhängig vom verwendeten Verfahren – die Klassenobjekte zusammengefaßt, die den geringsten Heterogenitätszuwachs (Verlust der Homogenität) aufweisen bzw. nach Fusion die homogenste neue Klasse bilden. Auf diese Weise ergibt sich eine Hierarchie von Klassen, die alle Objekte der Untersuchungseinheit einschließt³⁹.

Nach $N-1$ Gruppierungsschritten sind sämtliche Beobachtungseinheiten in einer Klasse zusammengefaßt. Eine übliche graphische Darstellungsweise dieses Klassifikationsverfahrens stellt das *Dendrogramm* dar.

Bock sieht Einsatzmöglichkeiten für hierarchische Verfahren darin, „daß sie einen Überblick über jene Objektgruppen liefert, die bei variierenden Homogenitätsforderungen sinnvollerweise als ‚Gruppe‘ angesehen bzw. eingerichtet werden können. Die Hierarchie zeigt, wie diese verschiedenen Gruppen miteinander zusammenhängen, ob sie durch sukzessive Fusion von Klassen vergleichbarer Größenordnung entstehen (intensive Gruppenstruktur) oder nur durch sukzessive Adjunktion benachbarter Einzelobjekte (schwache Gruppenstruktur)“⁴⁰. Hierarchische agglomerative Gruppierungsverfahren erzeugen somit eine ganze Folge von Klassen durch

sung linearer Zusammenhänge beschränkt. Schlosser, O., *Einführung in die sozialwissenschaftliche Zusammenhangsanalyse*, Reinbek 1976, S. 205 ff.

³⁶ Sodeur, *Empirische Verfahren*, S. 18.

³⁷ Sodeur, *Empirische Verfahren*, S. 21.

³⁸ Vgl. etwa Bock, *Automatische Klassifikation*, S. 316–355, S. 411–418.

³⁹ Vgl. Bock, *Automatische Klassifikation*, S. 360.

⁴⁰ Bock, *Automatische Klassifikation*, S. 358.

stufenweises optimales Zusammenfassen von Gruppen bzw. Objekten zu neuen Clustern – ein einmal konstruiertes Cluster kann in einem späteren Schritt nicht wieder aufgelöst werden –, aus denen die geeignete Gruppierung noch ausgewählt werden muß. Aufgrund dieser Eigenschaften sind die Einsatzmöglichkeiten hierarchisch-agglomerativer Verfahren beträchtlich eingeengt, beziehen sich doch die verschiedenen Fusionsverfahren nur auf „Teilaspekte der Klasseneinteilung, wie die der bestmöglichen Zusammenfassung zweier Elemente oder Klassen zu einer neuen Klasse oder auf die Homogenität einzelner Klassen“⁴¹; die gesuchten Gruppierungen selbst bleiben außerhalb.

Steinhausen-Langer beschreiben 8 verschiedene hierarchische Klassifikationsverfahren⁴², Verfahren, die hinsichtlich ihrer Gruppierungsergebnisse recht unterschiedlich sind. Folgt man der Bewertung der verschiedenen Verfahren bei Steinhausen-Langer und Bock⁴³, so sind vor allem das Verfahren *Average-Linkage* und das Verfahren *Ward* als der anstehenden Problemstellung angemessen zu bewerten. Anders als die anderen hierarchischen Verfahren verwenden die genannten Verfahren Heterogenitätsmaße im weiteren Sinne als Fusionskriterien. Diese Tatsache läßt übrigens auch die Verwendung der quadrierten Euklidischen Distanz als ein sinnvolles Ähnlichkeitsmaß erscheinen; das Verfahren *Ward* setzt dieses Distanzmaß aufgrund der Zielfunktion (Minimierung der Summe der quadrierten Euklidischen Abstände) voraus⁴⁴.

Während das Verfahren *Average-Linkage* die Ähnlichkeit von Gruppen über ein Distanzmaß feststellt, ist das Vorgehen im Verfahren *Ward* dadurch bestimmt, daß auf jeder Fusionsstufe ein Heterogenitätsmaß optimiert wird.

Auf jeder Stufe des Klassifikationsprozesses werden diejenigen Klassen/Objekte zusammengefaßt, für die der geringste Zuwachs zur Summe der Abstandsquadratsumme entsteht⁴⁵. Damit führt das Verfahren *Ward* zu dem gleichen Zielkriterium wie einige nicht-hierarchische Verfahren⁴⁶, allerdings mit dem Unterschied, daß das Verfahren *Ward* nur für jeden Fusionsschritt das Kriterium zu optimieren sucht, die nicht-hierarchischen Verfahren dagegen bei vorgegebener Klassenzahl direkt optimieren.

Die nicht-hierarchischen Klassifikationsverfahren dagegen teilen eine Gesamtheit von Beobachtungseinheiten in eine vorgegebene Anzahl von Klassen auf. Dies geschieht in der Regel *iterativ*⁴⁷. Verfahren dieses Typs zeichnen sich dadurch aus,

⁴¹ Sodeur, *Empirische Verfahren*, S. 143–144; vgl. auch Steinhausen-Langer, *Clusteranalyse*, S. 74 ff.

⁴² Steinhausen, D. und Langer, K., *Clusteranalyse, Einführung in Methoden und Verfahren der automatischen Klassifikation*, Berlin, New York 1977, S. 76–81.

⁴³ Steinhausen, Langer, *Clusteranalyse*, S. 76–81; Bock, *Automatische Klassifikation*, S. 387 ff.

⁴⁴ Vgl. Späth, H., *Cluster-Analyse-Algorithmen zur Objektklassifizierung und Datenreduktion*, München 1975, S. 171 ff.

⁴⁵ Vgl. Bock, *Automatische Klassifikation*, S. 407.

⁴⁶ Vgl. unten die Darstellung der nicht-hierarchischen Verfahren.

⁴⁷ Zu den nicht-iterativen nicht-hierarchischen Verfahren vgl.: Bock, *Automatische Klassifi-*

daß ein gegebenes globales Gütekriterium – bei vorgegebener Klassenzahl – durch Verschieben der Objekte so verbessert wird, daß ein – zumindest lokales – Optimum erreicht wird. Anders als bei hierarchischen Verfahren ist bei diesen Verfahren das Ergebnis des iterativen Prozesses die gesuchte Gruppierungszahl selbst. Ferner verbleiben einmal zugeordnete Objekte nicht wie bei den hierarchischen Verfahren immer in ihren Clustern, sondern werden bei vorgegebener Anfangspartition in einer Weise iterativ umgruppiert, daß eine Verbesserung des Gütekriteriums eintritt, bis sich auf diese Weise eine weitere Verbesserung nicht mehr erreichen läßt.

Zur nicht-hierarchischen iterativen Klassifikation gehört eine umfangreiche Gruppe verschiedener Verfahren, die sich vor allem unterscheiden in bezug auf 1. die Art der Erstellung der Anfangspartition, 2. in dem zu optimierenden Gütekriterium, 3. in der Anzahl gleichzeitig umzugrupprierender Objekte⁴⁸.

In der vorliegenden Untersuchung ist das erste Problem in folgender Form angegangen worden: 1. Es wird eine in einem hierarchischen Verfahren erzeugte Gruppierung als Anfangspartition vorgegeben. 2. Es wird eine durch die Siedlungsstruktur und die historisch-rechtliche Tradition gegebene Gruppierung als Anfangspartition vorgegeben. 3. Es wird eine Anfangspartition vorgegeben, die in folgender Weise generiert worden ist:

$\text{mod}(i - 1, K) + 1$, wobei $\text{mod}(a, b)$ = zahlentheoretische Funktion: Rest bei der Division ganzer Zahlen⁴⁹. Die Regel für die iterative Zuordnung der Objekte der Anfangspartition zu anderen Gruppen ergibt sich aus der Anwendung des *Minimal-Distanz-Verfahrens* und anschließendem *Austauschverfahren*⁵⁰.

Die Wahl der Zielfunktion ist nun von entscheidender Bedeutung für den Erfolg des Verfahrens, gehen doch über die Zielfunktion eine Reihe von Vorstellungen über Form und Größe der Gruppen in die Gruppierung selbst ein.

Die gebräuchlichsten Homogenitätskriterien sind das *Varianzkriterium* (= *Abstandsquadratsummenkriterium*, *Spur W-Kriterium*) und das *Determinantenkriterium*⁵¹. An dieser Stelle sollen noch einige Aspekte zur kritischen Auswahl der Kriterien angeführt werden. Die Verwendung des Varianzkriteriums als Zielfunktion ist nur dann sinnvoll, wenn 1. die Merkmale voneinander unabhängig sind, 2. die Klassen als *Kugeln* mit gleichem Radius im m -dimensionalen Merkmalsraum erscheinen, sowie 3. eine gleichmäßige Verteilung der Beobachtungseinheiten auf die Klassen zu erwarten ist⁵².

Diese Modellannahmen dürften als in der Realität höchst selten zutreffend be-
kation, S. 298 ff.

⁴⁸ Vgl. zu den iterativen nicht-hierarchischen Verfahren (= *partitionierende Verfahren*): Steinhausen, Langer, *Clusteranalyse*, S. 100–138; Bock, *Automatische Klassifikation*, S. 219 ff.; Späth, *Algorithmen*, S. 35–146.

⁴⁹ Vgl. Steinhausen, Langer, *Clusteranalyse*, S. 109.

⁵⁰ Vgl. Steinhausen, Langer, *Clusteranalyse*, S. 107–109, S. 118–119.

⁵¹ Vgl. Steinhausen, Langer, *Clusteranalyse*, S. 101–106; Bock, *Automatische Klassifikation*, S. 115, S. 162–195; Späth, *Algorithmen*, S. 58 ff.

⁵² Bock, *Automatische Klassifikation*, S. 162–163; Steinhausen, Langer, *Clusteranalyse*, S. 101–102.

wertet werden. Für diesen Fall wird demnach die Verwendung des Varianzkriteriums einen erheblichen Einfluß auf das Klassifikationsergebnis ausüben, es wird zu einer „künstlichen Klassifikation der Objekte führen, welche die wahre Gruppenstruktur der Daten nur verzerrt wiedergibt“⁵³. Das Varianzkriterium greift letztlich auf die *Euklidische Distanz* zurück, damit bestehen gegen die Verwendung dieses Kriteriums die gleichen Bedenken und Einwände wie gegen die Euklidische Distanz: Die Euklidische Distanz ist nicht invariant gegenüber nichtsingulären Transformationen, z. B. Skalentransformationen. Die fehlende Skaleninvarianz der Euklidischen Distanz ist dadurch zu umgehen, daß die Merkmale standardisiert werden, dabei ist die sog. Z-Standardisierung die am häufigsten angewendete Methode, deren Verwendung allerdings, wie Schlosser⁵⁴ deutlich gezeigt hat, nicht unproblematisch ist.

Beim Determinantenkriterium werden einige Voraussetzungen des Varianz-Kriteriums abgeschwächt: 1. Die Variablen sind voneinander abhängig, die Abhängigkeit ist in allen Gruppen gleich; 2. die zu den Klassen gehörenden Objekte bilden im Merkmalsraum Ellipsoide mit parallel ausgerichteten Achsen⁵⁵. Die Berechnung des Determinantenkriteriums ist im allgemeinen recht aufwendig; ferner nur dann sinnvoll, wenn $m < n-k$ (m = Zahl der Variablen, n = Zahl der Beobachtungseinheiten, k = Zahl der Klassen⁵⁶). Trotz der bekannten Vorteile des Determinantenkriteriums ist seine Einsatzmöglichkeit wegen des nicht unerheblichen Rechenaufwandes nur begrenzt. In dieser Untersuchung wird deshalb auch auf seine Anwendung in den nicht-hierarchischen Gruppierungsverfahren verzichtet.

Vergleicht man abschließend die hierarchischen und nicht-hierarchischen Verfahren, so ist festzustellen, daß die hierarchischen Verfahren aufgrund der Möglichkeit der Dendrogramm-Konstruktion ein wesentlich anschaulicheres Bild zur Struktur des Datensatzes bieten als die iterativen, nicht-hierarchischen Verfahren können. Letztere vermitteln lediglich eine Reihe von verschiedenen Gruppierungen, wobei es oft schwerfällt, die Beziehungen zwischen diesen Gruppierungen aufzudecken. Der besseren Anschaulichkeit hierarchischer Verfahren steht aber ein schwerwiegender Mangel gegenüber: Iterative Verfahren verfolgen durch ein wiederholtes *Verschieben* von Beobachtungseinheiten eine Zuordnung zu Gruppen zu erreichen, die ein Gütekriterium optimiert. Auf diese Weise ist es über die iterativen Verfahren im Gegensatz zu den hierarchischen Verfahren möglich, Gruppierungen, die sich schlechter in bezug auf das Gütekriterium erweisen, schrittweise zu korrigieren, eine Möglichkeit, die bei hierarchischen Verfahren nicht gegeben ist. Eine sich im späteren Verlauf des Gruppierungsprozesses als schlecht erweisende Zuordnung

⁵³ Vgl. Bock, *Automatische Klassifikation*, S. 163.

⁵⁴ Schlosser, *Zusammenhangsanalyse*, S. 56 ff.

⁵⁵ Vgl. Bock, *Automatische Klassifikation*, S. 186.

⁵⁶ Vgl. Späth, *Algorithmen*, S. 120.

läßt sich bei den hierarchischen Verfahren nicht korrigieren⁵⁷. Ein weiterer Vorteil nicht-hierarchischer Gruppierungsverfahren ist in der klar definierten Zielfunktion (*Spur-W-Kriterium/Determinantenkriterium*) gegeben.

3.2.2 Clusteranalyse und sozial-ökonomische Struktur

Mit Hilfe von Klassifikationsverfahren soll das Problem der Erfassung räumlicher Differenzen in der sozial-ökonomischen Struktur angegangen werden. Einander hinsichtlich ihrer Ausstattung ähnliche Gemeinden sollen dabei zu Gruppen zusammengefaßt werden, um auf diese Weise über eine Beschreibung der verschiedenen *Gemeinde-Gruppen* räumliche Unterschiede in der sozialen wie ökonomischen Struktur der Gemeinden und Ausstattungstypen zu ermitteln.

Das methodische Vorgehen wird sich dabei in folgender Form gestalten:

1. Es ist davon auszugehen, daß die Ergebnisse von statistischen Analysen in erheblichem Maße von sog. *Ausreißern* beeinflusst werden. Darum werden zunächst die 94 Gemeinden einem Test auf Ausreißer auf der Basis einer hierarchischen Klassifikation nach dem *Single-Linkage*-Verfahren unterzogen; die sich dabei ergebenden Ausreißer werden von der weiteren Untersuchung ausgeschlossen.
2. Von nicht zu unterschätzendem Einfluß auf das Ergebnis statistischer Analysen ist die Auswahl der Variablen. Die Zusammenstellung der Merkmale entsprechend den Untersuchungszielen ist bereits oben dargestellt worden, dabei bedürfen aber zwei Problemkreise noch einer besonderen Erörterung.

Zunächst soll das Problem der internen Gewichtung angesprochen werden, es sind zwei Apskete zu analysieren:

- (1) der Gesichtspunkt einer (inhaltlichen) *Redundanz* der Merkmale und
- (2) der Gesichtspunkt eines (statistisch-) *korrelativen Zusammenhangs* der Variablen.

Inhaltliche Redundanz besteht dann, wenn einige Variablen ähnliche oder zum Teil gleiche Eigenschaften der Objekte beschreiben, und führt folglich zu einer (unbeabsichtigten) internen Gewichtung der entsprechenden Merkmale. Mit dem Problem der inhaltlichen Redundanz eng verbunden ist das Problem statistisch-korrelativer Zusammenhänge, das ebenfalls zu einer Gewichtung bestimmter Merkmalsgruppen führt. Während im Falle inhaltlicher Redundanz Verzerrungen nur insoweit vermieden werden können, als es gelingt, „jeder gemeinten Eigenschaft genau ein beobachtetes Merkmal zuzuordnen (und, wenn nicht möglich) unter inhaltlich begründeten Vorentscheidungen (diese) . . . durch Indexbildung oder Skalierung (zusammenzufassen)“⁵⁸, ergeben sich im Falle (statistischer) Zusammenhänge Möglichkeiten, durch die Verwendung der Mahalanobis-Distanz⁵⁹ oder durch Faktorisierung in linear unabhängige Faktorenwerte interne Gewichtungen zu vermeiden.

⁵⁷ Vgl. Forst, *Klassifizierung von Städten*, S. 46

⁵⁸ Sodeur, *Empirische Verfahren*, S. 58–59.

⁵⁹ Zur Problematik der Verwendung der Mahalanobis-Distanz vgl. etwa: Steinhausen, Langer, *Clusteranalyse*, S. 59 ff.; Bock, *Automatische Klassifikation*, S. 40 ff., S. 84 ff.

Beide Verfahren sind jedoch mit gravierenden Nachteilen behaftet. Durch dieses Vorgehen „werden . . . datenimmanente bzw. stichprobenimmanente Linearität kontrolliert und damit für die weitere Analyse ausgeschaltet“⁶⁰. Diese Zusammenhangsinformationen können daher im Klassifikationsverfahren keine Berücksichtigung finden, obwohl doch Abhängigkeiten linearer wie nicht-linearer Art zwischen den Merkmalen eine notwendige Voraussetzung für die Bildung nicht kontinuierlicher Verteilungsmuster im Merkmalsraum sind⁶¹.

Die Ausschaltung statistisch-korrelativer Zusammenhänge führt also dazu, daß für die Klassifikation der Objekte wesentliche Informationen nicht beachtet würden.

Die beiden Aspekte, inhaltliche Redundanz wie statistisch-korrelative Zusammenhänge, lassen sich mit Hilfe statistischer Methoden nicht trennen. Es ist darum auch nicht unbedingt sinnvoll, über spezielle Abstandsmaße bzw. Zielfunktionen eventuelle Korrelationen ausschalten zu wollen, vielmehr erscheint es der speziellen Zielsetzung der Clusteranalyse, latente Strukturen in Objektmengen unter einer speziellen Fragestellung aufdecken zu wollen, angemessen zu sein, wenn die Merkmale unter theoretischen Gesichtspunkten gemäß der Problemstellung ausgewählt werden und die dabei entstehenden Korrelationen quasi als erwünschte interne Gewichtung von Teilaspekten angesehen werden.

Ein anders gelagertes Problem stellt die Frage nach der Reduktion des Merkmalsraumes dar. Nach Sodeur stellt sich das Problem wie folgt: „Die Elemente (sollen) statt im ursprünglichen Merkmalsraum unter möglichst vollständigem Erhalt der Struktur in einen neuen Raum mit möglichst geringer Dimensionszahl dargestellt werden“⁶². Es handelt sich hier um eine rein technische Reduktion des Merkmalsraumes. „Keineswegs wird damit das . . . Problem schiefwinkliger Merkmalsachsen durch ‚Orthogonalisierung‘ gelöst. Vielmehr wird durch die gegebene Verteilungsstruktur der Elemente, sei sie nun verzerrt oder nicht, bei der Reduktion des Merkmalsraumes weitestmöglich erhalten“⁶³.

Die Hauptkomponentenanalyse stellt auf der Ebene metrischer Daten in diesem Sinne ein übliches Verfahren dar. Ihre Anwendung wird in der Literatur jedoch kontrovers diskutiert. Während z. B. Steinhausen und Langer⁶⁴ in der Hauptkomponentenanalyse ein sinnvolles Verfahren der Reduktion des Merkmalsraumes sehen, gilt

⁶⁰ Schlosser, *Zusammenhangsanalyse*, S. 140.

⁶¹ Vgl. hierzu auch: Schlosser, *Zusammenhangsanalyse*, S. 25 ff., S. 140.

⁶² Sodeur, *Empirische Verfahren*, S. 59.

⁶³ Sodeur, *Empirische Verfahren*, S. 61. Zur Hauptkomponententransformation sowie die sich in diesem Zusammenhang ergebenden Schwierigkeiten einer inhaltlichen Interpretation vgl.: Überla, *Faktorenanalyse*, S. 240; Bock, *Automatische Klassifikation*, S. 237 ff.

⁶⁴ Steinhausen, Langer, *Clusteranalyse*, S. 129, S. 176 oder: Steinhausen, D., *Clusteranalyse großer Datenmengen – Effektive Verfahren und Programme zur Clusterung qualitativer Daten*, Münster 1977, S. 8 ff., S. 18 (= Rechenzentrum Universität Münster, Schriftenreihe Nr. 18).

dieses Verfahren, wenn man den Ausführungen Schlossers⁶⁵ folgt, als nicht empfehlenswert: Besondere Probleme ergeben sich u. a.

(1) aus den dem Verfahren zugrunde liegenden Zusammenhangsbegriff, der impliziert, daß alle linearen Zusammenhangsinformationen ausgefiltert werden,

(2) aus der bei (linearer) Projektion gegebenen Veränderung der relativen Lage der Punktkonfiguration, die dazu führen kann, daß die im ursprünglichen Merkmalsraum vorhandene Gruppierung anhand der projizierten Punkte nicht mehr erkannt werden kann, da verschiedene Gruppen zusammenfallen: Für den Fall, daß die Klassenzahl größer als die Zahl der Variablen ist, kann ein solcher Fall nicht ohne weiteres ausgeschlossen werden⁶⁶.

3. Wegen der Problematik sowohl einer Reduktion als auch einer Transformation des Variablenraumes über eine Hauptkomponententransformation erfolgt eine Klassifikation der Gemeinden auf der Basis des ursprünglichen Merkmalsraumes nach vorheriger z-Transformation. Die quadrierte Euklidische Distanz dient dabei als Ähnlichkeitsmaß. Über ein hierarchisches Verfahren soll ein Überblick über die sozial-ökonomische Struktur der Gemeinden gewonnen werden.

4. Ausgehend von der Gruppierung auf der Grundlage eines hierarchischen Clusterverfahrens erfolgt anschließend über iterative nicht-hierarchische Verfahren eine endgültige Fixierung der Gruppierungsstruktur.

Die gefundenen Cluster werden mit Hilfe der Cluster-Mittelwerte im Vergleich zum Gesamtmittelwert charakterisiert.

3.2.2.1 Analyse auf Ausreißer

Unter *Ausreißern* sollen solche Beobachtungen oder Gruppen verstanden werden, die deutlich von den zentralen multivariaten Punkthäufungen abgetrennt sind, d. h. die unter geometrischen Gesichtspunkten im Vergleich zu anderen Beobachtungen weit von Häufungen von Beobachtungen entfernt liegen. Neben diesen sog. *multivariaten* Ausreißern existieren selbstverständlich auch *eindimensionale* Ausreißer. Beide Formen haben eine nicht unbedeutende, im allgemeinen nicht erwünschte Bedeutung für die Gruppierung. Die verschiedenen Klassifikationsverfahren sind nämlich höchst unterschiedlich in bezug auf die Verarbeitung von Ausreißern.

Das *Single-Linkage*-Verfahren ist wohl die bekannteste Methode, die entferntere Objekte bevorzugt und sie einer Gruppe zuordnet⁶⁷. Aber auch die nicht-hierarchischen Verfahren haben die Tendenz, Ausreißer Gruppen zuzuordnen, die von ihnen deutlich getrennt sind.

⁶⁵ Schlosser, *Zusammenhangsanalyse*, S. 205 ff.; vgl. hierzu auch: Bock, *Automatische Klassifikation*, S. 241 ff.

⁶⁶ Vgl. Bock, *Automatische Klassifikation*, S. 242.

⁶⁷ Vgl. Steinhausen, Langer, *Clusteranalyse*, S. 75; vgl. auch: Bock, *Automatische Klassifikation*, S. 390–391. Der für das *Single-Linkage*-Verfahren typische Aneinanderreihungseffekt wird übrigens in dem unten zu besprechenden Test zur Erkennung von Ausreißern genutzt.

Die Ausschaltung von Ausreißern ist aber nicht nur deshalb sinnvoll, weil sie den Klassifikationsprozeß deutlich stören können, sondern auch, weil sie einen verfälschenden Einfluß auf die Korrelationsbeziehungen zwischen den Merkmalen ausüben. Ein ebenso gewichtiges Argument für die Beseitigung von Ausreißern ist, daß es weder sinnvoll ist, Ausreißer bestimmten Gruppen zuzuordnen, noch sie als eigenständige Gruppen zu betrachten.

Der Test auf multivariate Ausreißer wird nach einem von F.J. Rahlf vorgeschlagenen, leicht durchzuführenden Verfahren⁶⁸ durchgeführt. Dabei geht man davon aus, daß die Ausreißer im mehrdimensionalen Merkmalsraum in einer gewissen Entfernung von der übrigen *Punktewolke* liegen. Die relative Lage der Ausreißer zu den anderen Objekten bestimmt man nun durch die Konstruktion eines sog. *Minimalbaumes*⁶⁹ der Verbindungen zwischen den einzelnen Objekten. Aus den bekannten Eigenschaften des Minimalbaumes ergibt sich, daß große Distanzen durch einen großen Abstand der durch die entsprechende Linie verbundenen Objekte verursacht werden. Es ist nun zu prüfen, in welchem Umfang sich die *großen* Distanzen von den übrigen Distanzen des Minimalbaumes signifikant unterscheiden.

Folgende Testgröße ist dazu zu berechnen:

$$G^* = \frac{\max(d_i^2)}{\bar{d_i^2}} \quad ; \text{ dabei sind}$$

$\max(d_i^2)$ die maximale quadrierte Euklidische Minimaldistanz

$\bar{d_i^2} (i=1, \dots, n-1)$ der Durchschnitt aller quadrierten Euklidischen Minimaldistanzen $\max(\bar{d_i^2})$;

d_i^2 sind näherungsweise gammaverteilt. Die Schätzung der Parameter α und ν erfolgt über Ranggrößen nach Wilk/Gnanadesikan/Huyett⁷⁰. G^* folgt unter der Nullhypothese unabhängig von α einer Beta-Verteilung, deren Parameter können allein mit Hilfe von ν berechnet werden. Bei Rahlf sind für unterschiedliches ν und n obere kritische Signifikanzgrenzen G bei unterschiedlichem Signifikanzniveau tabelliert.

Der mit Hilfe des *Single-Linkage*-Verfahren erzeugte Stammbaum zeigt deutlich, daß die Städte Münster, Warendorf und Telgte im Vergleich zu den anderen Gemeinden relativ große Distanzen zu den übrigen Objekten bzw. Gruppen haben. Der Ausreißer-Test für diese Städte führt dabei zu dem folgenden Ergebnis, daß alle drei Städte auf einem Signifikanzniveau von $\alpha = 0,10$ als Ausreißer zu betrachten sind.

⁶⁸ Rohlfs, F.J., *Generalisation of the gap test for the detection of multivariate Outliers*, in: *Biometrics*, 31 (1975), S. 93–101.

⁶⁹ Steinhausen, Langer, *Clusteranalyse*, S. 94–97.

⁷⁰ Wilk, M.B., Gnanadesikan, R., Huyett, J.J., *Estimation of parameters of the fama distribution using order statistic*, in: *Biometrika*, 49 (1962), S. 525–535.

3.2.2.2 Gruppierung unter Einschluß der Ausreißer

Nach der Feststellung der störenden Ausreißer soll nun im vollständigen und reduzierten Datensatz in einem zweistufigen Verfahren zunächst hierarchisch und anschließend iterativ, auf den Ergebnissen der hierarchischen Gruppierung aufbauend, klassifiziert werden.

Zunächst wird die mit dem vollen Datensatz durchgeführte Analyse dargestellt. Bei den zwei durchgeführten hierarchischen Klassifikationen stellt sich heraus, daß das *Ward*-Verfahren die am besten zu interpretierenden Ergebnisse liefert, darum wird auf eine Darstellung der Ergebnisse nach *Average Linkage* verzichtet.

Einen ersten, sehr informativen Überblick über die vorhandene Struktur liefert das Dendrogramm nach dem *Ward*-Verfahren. Der Stammbaum zeigt, daß die Untersuchungsgesamtheit in zwei deutlich voneinander getrennte Gruppen zerfällt. Die eine Gruppe wird von den relativ großen Landstädten (Ahlen, Beckum, Lüdinghausen, Telgte, Warendorf, Werne) und zwei Minderstädten (Oelde, Olfen) und der Stadt Münster gebildet. Zur anderen Gruppe gehören die Gemeinden des flachen Landes einschließlich der restlichen Minderstädte. Anhand des Dendrogramms ist es möglich, eine feinere Untergliederung vorzunehmen. So läßt sich die Zahl der Gruppen, die sinnvoll zu bilden sind, einigermaßen gut abschätzen. Nach dem vorliegenden Bild erscheinen 5 Gruppen als empfehlenswert. Dabei ist nun bemerkenswert, daß die Städte Münster, Telgte und Warendorf (Gruppe 1, 3, 5) je eine eigene Gruppe bilden. Eine gemeinsame Gruppe bilden Ahlen, Beckum, Oelde, Lüdinghausen, Werne und Olfen (Gruppe 4). Die Landgemeinden, die kleineren Städte und Minderstädte werden einer Gruppe mit 85 Mitgliedern zugeordnet.

Bevor die Ergebnisse der Gruppierung nach dem Verfahren Ward einschließlich der Versuche zur Verbesserung der Ergebnisse mit Hilfe eines iterativen Verfahrens genauer dargestellt werden, soll kurz auf einen Versuch eingegangen werden, die Untersuchungsgesamtheit auf der Grundlage unterschiedlicher rechtlicher Traditionen, wie siedlungstopographischer Gegebenheiten zu gruppieren. Die folgenden Kategorien sind unterschieden worden: Stadt, Minderstadt, Kirchdorf und Kirchspiel. Die sich auf dieser Grundlage ergebende Gruppierung der Untersuchungsgesamtheit ist mit Hilfe des iterativen Minimaldistanz-Verfahrens und dem nachgeschalteten Austauschverfahren verbessert worden⁷¹. Die Ergebnisse dieser Analyse werden in einer Kontingenztafel der Ausgangspartition gegenübergestellt.

⁷¹ Varianz-Kriterium = 17,22.

⁷² Varianz-Kriterium = 10,63.

Tabelle 5
Kontingenztafel für die Gruppenzugehörigkeit

<i>Gruppen-Nr. bei Minimal-Distanz- und Austauschverfahren</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	
<i>Gruppen-Nr. bei Ausgangspartition</i>					
1 (Stadt)	4	2	0	0	6
2 (Minderstadt)	3	0	0	7	10
3 (Kirchdorf)	0	0	29	30	59
4 (Kirchspiel)	0	0	19	0	19
	7	2	48	37	94

Als Maß des Zusammenhangs wird das symmetrische Kontingenzmaß λ bestimmt. Es nimmt im vorliegenden Fall mit $\lambda = 0,22$ einen recht geringen Wert an. Dieser geringe Wert ist vor allem darauf zurückzuführen, daß die Gruppe 3 (Kirchdorf) im iterativen Verfahren auf 2 Gruppen (Gruppe 3 und 4) aufgeteilt wird. Eine Aufspaltung ist auch bei den Gruppen 1 (Stadt) und 2 (Minderstadt) festzustellen. Der geringe Wert für λ zeigt also recht deutlich, daß zwischen beiden Gruppierungen so gut wie kein Zusammenhang besteht; die Gruppierung der Untersuchungsgesamtheit nach rechtshistorischen bzw. siedlungsgeographischen Aspekten spiegelt damit also in keiner Weise die Gliederung der Gemeinden nach sozial-ökonomischen Strukturmerkmalen wider.

Zur endgültigen Festlegung der Gruppenanzahl werden die Werte des Varianzkriteriums für verschiedene Gruppenzahlen sowohl für eine programmintern erzeugte Anfangspartition als auch für eine sich aus der Anwendung des Ward-Verfahrens ergebende Gruppierung errechnet.

Die Werte des Varianzkriteriums sind in der *Tabelle 6* zusammengestellt, sie zeigen sowohl für die programminterne Gruppierung als auch für die Gruppierung nach Ward einen relativ gleichmäßig sinkenden Verlauf bei steigender Gruppenzahl, aller-

Tabelle 6
Varianzkriterium in Abhängigkeit von der Gruppenanzahl ($N = 94$)

<i>Gruppenanzahl</i>	<i>programminterne Partition</i>	<i>Partition nach Ward</i>
2	23,12	23,12
3	21,58	21,83
4	17,26	20,80
5	18,35	19,52
6	16,44	18,39
7	14,50	17,41
8	14,34	16,81
9	15,48	16,20
10	14,87	15,79
11	15,77	15,63
12	14,09	15,28
13	15,12	15,08

dings erreichen die Werte des Varianzkriteriums bei programminterner Partition bei den Gruppenzahlen 5, 9 und 11 ein relatives Maximum; das absolute Maximum liegt bei 2 Gruppen.

Es zeigt sich jedoch, daß in dem analysierten Bereich von 2 bis 13 Gruppen die Klassifikationsergebnisse bei einer durch das hierarchische Verfahren *Ward* vorgegebenen Anfangspartition deutlich besser sind als die Ergebnisse bei Standardpartition.

Wird auf der Grundlage des varianzanalytischen Kriteriums eine Bestimmung der optimalen Klassenanzahl über das absolute Maximum bzw. relative Maxima vorgegeben, so ergeben sich im Bereich 2 bis 13 Klassen, ein absolutes Maximum bei 2 und ein relatives Maximum bei 11 Klassen. Da eine differenzierte Beschreibung der Untersuchungsgemeinden auf der Grundlage einer 2-Klassenlösung nicht sinnvoll erscheint, scheidet diese Lösung aus. Auch die über das Varianzkriterium ermittelte 11-Gruppenlösung ist wenig aussagekräftig, weil auf diesem Gruppierungsniveau schon zu stark individuelle Strukturausprägungen in den Untersuchungseinheiten Gewichtung finden, die allgemeinen Strukturierungsleitlinien dagegen nicht sichtbar werden. Aus diesem Grunde scheidet auch die Partition mit 11 Gruppen für eine weitere Analyse aus.

Versucht man dennoch, das Varianzkriterium für die Gruppenbildung zu berücksichtigen, so bietet sich die schon nach dem Dendrogramm des *Ward*-Verfahrens abzeichnende Lösung mit 5 Gruppen an, da das Varianzkriterium bei Standardpartition ein relatives Maximum an dieser Stelle aufweist. Zur *ersten* Gruppe gehört die Stadt Münster, zur *zweiten* Telgte und zur *dritten* Ahlen, Oelde, Freckenhorst Std., Beckum Std., Olfen Std., Werne Std. und Lüdinghausen Std., zur *vierten* Warendorf und die *fünfte* Gruppe umfaßt die restlichen 84 Gemeinden. An dieser Stelle genügt es, die Gruppen 1, 2 und 4 näher zu charakterisieren, da die Mitglieder dieser Gruppe als Ausreißer bei weiteren Gruppierungsversuchen ausgeschlossen bleiben.

Innerhalb der Gesamtheit aller Gemeinden nimmt Münster eine Sonderstellung ein, in der sich die besondere geschichtliche Position der ehemaligen Landeshauptstadt widerspiegelt. Im Dienstleistungs- und Gewerbesektor besteht über eine weit überdurchschnittliche Versorgung im Bereich des täglichen Bedarfs ein weit gestreutes Angebot mit Gütern und Diensten des periodischen Bedarfs. Der hohe Versorgungsgrad in der Grund- wie der Ergänzungsausstattung macht die Bedeutung der Stadt als regionales Versorgungszentrum deutlich.

Wie Münster verfügt auch Warendorf über eine breit gefächerte Ausstattung im gewerblichen und Dienstleistungsbereich mit zahlreichen hochspezialisierten Sparten. Dabei tritt das Textil- und Bekleidungsgewerbe mit einer stark ausgeprägten Vielfalt hervor.

Telgte verfügt über eine durchschnittliche Grundausrüstung; einige Bereiche des Dienstleistungssektors weisen aber eine überdurchschnittlich hohe Besetzung auf. Dabei handelt es sich um Metallwaren-, Wein- und Getreide- sowie insbesondere Handel mit Bildern; letzterer ist auf die Funktion der Stadt Telgte als Wallfahrtsort zurückzuführen.

3.2.2.3 Gruppierung nach Ausschaltung der Ausreißer

Die aufgrund des multivariaten Ausreißertests als solche identifizierten Gemeinden, Münster, Telgte und Warendorf werden von der weiteren Analyse ausgeschlossen. Die verbleibenden 91 Gemeinden werden wiederum zunächst mit Hilfe des hierarchischen Ward-Verfahrens gruppiert, die Ergebnisse anschließend im iterativen Verfahren verbessert. Parallel dazu wird eine iterative nicht-hierarchische Gruppierung mit einer programminternen Startpartition durchgeführt.

Wie schon bei der Klassifikation auf der Basis der Gesamtheit der Raumeinheiten soll nun zunächst eine Überprüfung der Gruppierung auf der Grundlage rechts-historisch-siedlungsgeographischer Kriterien mit Hilfe des nicht-hierarchischen Verfahrens (*Minimaldistanz-, Austauschungsverfahren*) erfolgen. Die Überprüfung der Stabilität dieser Gruppierung an Hand des Kontingenzmaßes λ ergibt für $\lambda = 0,29$ einen Wert, der nur geringfügig über dem oben bestimmten Maß liegt. Auch in diesem Fall wird das geringe Kontingenzmaß im wesentlichen bedingt durch die Aufspaltung der Gruppe 3 (Kirchdorf) auf zwei Klassen. Auf mehrere Klassen wird ferner ebenfalls wieder die Gruppe 2 (Minderstadt) aufgeteilt. Damit wird in einer um die Ausreißer verkürzten Gesamtheit das in der Analyse des Gesamttraumes erzielte Ergebnis bestätigt. Eine Übereinstimmung zwischen den traditionellen Siedlungskategorien und einer Gruppierung nach sozialen und ökonomischen Merkmalen besteht nicht⁷³.

⁷³ Zur Gruppe 1 gehören:

St. Mauritz	Füchtorf	Sendenhorst Std.
Lamberti	Sassenberg	Sendenhorst Ksp.
Überwasser	Harsewinkel Ksp.	Drensteinfurt Std.
Handorf	Marienfeld	Drensteinfurt Ksp.
Amelsbüren	Greffen	Herbern
Hiltrup	Beelen	Bockum
Telgte	Freckenhorst Ksp.	Hövel
Westbevern	Everswinkel	Walstedde
Wolbeck Std.	Hoetmar	Werne Ksp.
Wolbeck Ksp.	Westkirchen	Stockum
Angelmodde	Ostenfelde	Kapelle
Alverskirchen	Ennigerloh	Südkirchen
Albersloh	Oelde Ksp.	Bork
Rinkerorde	Stromberg	Altünen
Nottuln	Wadersloh	Lüdinghausen Ksp.
Roxel	Liesborn	Senden
Albachten	Diestedde	Venne/Ottmarsbocholt
Bösensell	Herzfeld	Ascheberg
Nienberge	Lippborg	Nordkirchen

Für den Bereich von 2 bis 10 Gruppen sind die Werte des Varianz-Kriteriums sowohl bei Vergabe der hierarchischen Klassifikation als auch bei Standardpartition errechnet worden:

Tabelle 7
Varianzkriterium in Abhängigkeit von der Gruppenanzahl (N = 91)

<i>Gruppenanzahl</i>	<i>Partition nach Ward</i>	<i>programminterne Partition</i>
2	17,58	17,72
3	12,77	12,54
4	10,87	10,76
5	10,13	9,20
6	9,60	9,40
7	9,37	8,29
8	9,31	8,55
9	9,27	8,68
10	9,21	8,59

Es zeigt sich, daß bei Vorgabe der Klassifikationsergebnisse nach *Ward* auch in diesem Fall wiederum die Gruppierungen besser sind als bei programmgesteuerter Anfangsgruppierung. Wie schon bei der Analyse des vollständigen Datensatzes ergeben sich aber auch Schwierigkeiten, die optimale Gruppenanzahl über ein relatives Maximum des Varianzkriteriums zu ermitteln. Unter Berücksichtigung des Verlaufs der beiden Reihen des Varianzkriteriums erscheint eine Gruppierung mit sechs Gruppen auf der Basis der iterativen Gruppierung bei einer Anfangspartition nach *Ward* sinnvoll.

Durch die Eliminierung der Ausreißer ist es zu einer weiteren Untergliederung der verbleibenden Städte gekommen: Olfen und Lüdinghausen bilden nun eine eigene Gruppe, die der aus Ahlen, Beckum und Oelde bestehenden Gruppe gegenüber-

Fortsetzung Anm. 73

Greven	Altahlen	Selm
Gimbte	Neuahlen	Olfen Ksp.
Saarbeck	Heessen	Seppenrade
Ladbergen	Dolberg	
Ostbevern	Lütke Untrop	
Milte	Beckum Ksp.	
Einen	Sünninghausen	
Lienen	Vellern	
Altwarendorf	Vorhelm	
Neuwarendorf	Enniger	

steht. Eine weitere Aufteilung erfährt auch die 84 Mitglieder große Gruppe 5; insgesamt bleibt aber trotzdem noch eine relativ große Gruppe (80 Mitglieder) bestehen.

Die *erste Gruppe*⁷³ besteht im wesentlichen aus Gemeinden, die von einer sog. *Durchschnittsgemeinde* die geringsten Abweichungen aufweisen. Innerhalb der Wirtschaftsstruktur lassen sich keinerlei Schwerpunkte feststellen. Diese Gruppe wird erst bei einer Gruppierungslösung mit 9 Gruppen in weitere Klassen aufgeteilt. Die Städte Drensteinfurt, Sendenhorst und Wolbeck bilden dann mit einer Reihe anderer Gemeinden⁷⁴ eine eigene Gruppe, die in ihrer wirtschaftlichen Struktur sich aufgrund einer besseren Grundausrüstung in den Wirtschaftsunterabteilungen Metallgewerbe, Holzverarbeitendes Gewerbe und Lederverarbeitung von den verbleibenden Gemeinden der Gruppe 1 unterscheiden.

Die *zweite Gruppe* (Beckum, Ahlen, Oelde) enthält die größeren Städte, die vor allem durch eine durchschnittliche bis überdurchschnittliche Grundausrüstung sämtlicher Wirtschaftsbereiche gekennzeichnet sind. Über diese Grundausrüstung hinaus besteht im Bereich des produzierenden Gewerbes eine relativ gute Ausstattung mit Gewerben, die Güter des gehobenen Bedarfs produzieren. In der Metallverarbeitung fallen die gute Versorgung mit feinmechanischen Betrieben, in der Papierverarbeitung mit Druckereibetrieben und Buchbindereien und schließlich in der Textilverarbeitung mit Hut- und Putzmachereibetrieben auf. Die Städte dieser Gruppe haben ferner ein gutes Angebot im Dienstleistungsbereich, im Handel besteht ein breit gefächertes Angebot, das Gaststättenwesen ist gut ausgebaut.

Die ebenfalls nur drei Gemeinden (Freckenhorst Std., Appelhülsen, Schapdetten) umfassende *dritte Gruppe* hat ähnlich wie die Gruppe 1 eine nur durchschnittliche Ausstattung in den verschiedenen Wirtschaftsabteilungen. Im Unterschied zur ersten Gruppe sind aber in den Bereichen Baugewerbe, Holzverarbeitung sowie Leder- und Textilverarbeitung Schwerpunkte zu erkennen.

Die eine Gemeinde der *vierten Gruppe* (Std. Harsewinkel) steht von ihrer Struktur her zwischen den Gemeinden der zweiten und der dritten Gruppe. Die Ausstattung im gewerblichen Sektor geht in verschiedenen Bereichen deutlich über die Grundausrüstung hinaus, so in relativ hoch spezialisierten Sparten der Metallverarbeitung, der Genußmittelherstellung. Nicht unwichtig ist auch die Landwirtschaft. Im Dienstleistungssektor verfügt diese Gruppe über ein recht vielseitiges Angebot im Bereich des Handels.

⁷⁴ Zur Gruppe 2 gehören:

Sassenberg	Herbern
Wadersloh	Südkirchen
Ennigerloh	Nottuln
Lippborg	Nordkirchen
Venne/Ottmarsbocholt	Stromberg
Greven	Hövel

Die *fünfte Gruppe* (Lüdinghausen Std., Olfen Std., Werne Std.) ähnelt von ihrer wirtschaftlichen Struktur her stark der Gruppe 2. In der Grundausrüstung bestehen zwischen den Gemeinden beider Gruppen nur graduelle Unterschiede. Allerdings haben sich in den Gemeinden des Typs 5 deutlich andere Schwerpunkte wirtschaftlicher Tätigkeit herausgebildet. Ein solcher Schwerpunkt ist etwa die Seifen- und Farbenherstellung, oder das Kürschnergewerbe.

Die *sechste Gruppe* (Havixbeck) verfügt in fast allen Bereichen des gewerblichen Sektors und des Dienstleistungssektors über eine nur durchschnittliche Ausstattung; sie ähnelt damit sehr stark der Gruppe 1. In einigen Bereichen, der Steinbearbeitung und Keramikherstellung, im Bauhandwerk, in einigen Sparten der Holzverarbeitung, aber auch im Großhandelsbereich, verfügt diese Gruppe über eine überdurchschnittliche Ausstattung. Auch dem agraren Sektor kommt in dieser Gruppe eine gewisse Bedeutung zu.

4. Zusammenfassung

Zur Erfassung und Beschreibung struktureller Unterschiede zwischen Stadt- und Landgemeinden werden die Faktorenanalyse und numerische Klassifikationsverfahren auf die Möglichkeiten des Einsatzes geprüft. Die Untersuchungen zeigen, daß eine sinnvolle Reduktion des Datenmaterials mit der Faktorenanalyse nicht erreicht werden kann. Ein repräsentativer Faktor zur Kennzeichnung der sozio-ökonomischen Struktur kann nicht abgeleitet werden, die Zusammenfassung der vielen Merkmale zu einer einzigen, die Ausstattung der Gemeinden kennzeichnenden Kennziffer ist nicht möglich. Trotz verschiedener Modifikationen des faktorenanalytischen Ansatzes — Kommunalitätenschätzung, schiefwinklige Rotation — ergeben sich keine sinnvoll zu interpretierende Faktoren, die als Grundlage zu weiteren Untersuchungen dienen könnten.

Der Einsatz numerischer Klassifikationsverfahren mit dem Ziel, die Gemeinden entsprechend ihrer Variablenausprägungen in möglichst homogene und voneinander gut zu trennende Klassen aufzuteilen, um auf diese Weise typische Strukturmuster aufzudecken, bringt, dies zeigen die Klassifikationsergebnisse, im Vergleich zur Faktorenanalyse wesentliche Fortschritte.

Die Gruppierung der Untersuchungsgesamtheit erfolgt in einem zweistufigen Verfahren: Auf der Basis des hierarchisch-agglomerativen Verfahrens nach Ward wird zur Verbesserung der Gruppierung ein iteratives Verfahren verwendet. Der gleichen Zielsetzung dient eine multivariate Ausreißeranalyse.

Die Gruppierung nach allen Merkmalen macht erhebliche Strukturunterschiede sichtbar, die sich aber nicht auf die Dichotomie Stadt-Land reduzieren lassen.

Summary: Factor Analysis and Cluster Analysis. An Example of Applikations of Multivariate Procedures in Analysing the Relationship between Town and Country in Eastern Münsterland in the Early 19th Century

By the end of the 18th century the model of division of labour between town and country — existant since the late Middle Ages — in which the country served above all to bring forth agricultural products and the town was the favoured station of trade and service, had been largely abolished though the exceptional legal position of the town generally still continued. The problem of the position of the towns in relation to the communities of the country was investigated especially from an economic viewpoint by exploring a district in the eastern Münsterland. The investigation was based on 55 socio-economical characteristics which were chosen out of the „statistisch-topographische Tabelle“ of 1816, an early statistics of Prussian communities. In order to describe the structural differences between town and country communities the factoranalysis and numerical procedures of classification were considered for possible use. The inquiries showed that a reasonable reduction of the data could not be achieved by factoranalysis. A representative factor could not be deduced for the characterization of the socio-economical structure. In spite of different modifications of the factoranalytical approach there were no factors which could be interpreted reasonably and which could serve as basis of further investigations. In comparison to the factoranalysis, significant improvement was achieved using the clusteranalysis to attempt a division of the communities into homogeneous clusters that could be easily separated from each other. This aim was pursued in order to disclose the typical structure patterns. The clustering ensues from a procedure done in two steps: An iterative procedure was used on the basis of Ward's hieratic-agglomerative procedure to improve the grouping. The grouping according to all characteristics made evident considerable structure differences, which nevertheless cannot be reduced to the town-country dichotomy.